Journal of Experimental Social Psychology 108 (2023) 104495

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Evidence for bidirectional causation between trait and mental state inferences☆

Chujun Lin *, Mark Thornton

Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755, United States of America

A R T I C L E I N F O

Keywords:
Person perception
Trait
Mental state
Causal relationship
Statistical learning

A B S T R A C T

To navigate the social world, people must understand each other's momentary thoughts and feelings (mental states) and enduring personalities (traits). How do people make trait and mental state inferences in the real world? Prior research has artificially separated these two processes, primarily studying each topic on its own. However, in real life, these two processes constantly co-occur and rely on partially overlapping information. It is likely that people inform one type of inferences with the other. Here we investigate this possibility using naturalistic paradigms, statistical learning, and stimulus optimization techniques. We first demonstrated the correlation between trait and mental state inferences of targets in naturalistic videos (Study 1) and familiar others in real life (Study 2). Targets perceived to have similar traits were judged to experience various mental states with similar frequencies. We showed that this association was causal in two experiments. Learning that two people experienced similar mental states across a range of situations caused participants to attribute similar traits to them (Study 3). Conversely, observing two people had similar traits caused participants to attribute similar mental states to them across a range of situations (Study 4). Together, these four preregistered studies (total $N =$ 762) reveal that trait and mental state inferences continually run in parallel and that people rely on others' enduring traits to predict their momentary states, and vice versa. These findings highlight that biases in trait impressions may distort understanding of others' situation-specific states, and others' uncharacteristic states may influence judgments of their enduring traits.

1. Introduction

People describe each other using a wealth of features, such as their personality traits, emotional states, attitudes, goals, and beliefs (Fiske, Cuddy, & Glick, 2007; Uleman, Adil Saribay, & Gonzalez, 2008). Understanding each one of these features independently helps people better navigate the complex social world. For instance, understanding others' emotions facilitates accurate evaluation of social situations, increases the quality of social and romantic relations, and enhances workplace performance (Lopes, Salovey, Côté, & Beers, 2005; Schutte et al., 2001; Van Rooy & Viswesvaran, 2004; Wu, Schulz, Frank, & Gweon, 2021). Understanding our friends' and family members' personalities guides social interactions, helping us to know that we should approach open-minded friends for feedback on new ideas, or emotionally stable family members for social support (Fleeson, 2001; Sherman, Rauthmann, Brown, Serfass, & Cooper, 2015; Stachl et al., 2020).

How do people obtain different types of social knowledge about others in the real world? Given the assumption that each type of social knowledge could function independently, the majority of prior research has studied each topic on its own without reference to the other. For instance, researchers interested in the theory of mind focus on how people infer others' mental states, and study this process using mainly momentary cues such as facial expressions and social scenarios (Adolphs, Mlodinow, & Barrett, 2019; Ekman & Friesen, 1971; Jamali et al., 2021; Thornton & Tamir, 2020). On the other hand, researchers interested in the topic of person perception focus on how people infer others' traits, and study this process using mainly trait-implying cues such as static facial structures and trait-implying behavior (Hackel, Mende-Siedlecki, & Amodio, 2020; Lin, Keles, & Adolphs, 2021; Stolier, Hehman, & Freeman, 2020; Todorov & Uleman, 2003).

However, in everyday experience, the processes of trait and mental state inferences constantly co-occur, and rely on partially overlapping information (Zaki, 2013). How people infer others' traits may be influenced by their inferences of others' mental states, and vice versa. Since

☆ This paper has been recommended for acceptance by Editor name: Shlomo Hareli

* Corresponding author.
E-mail address: Chujun.Lin@Dartmouth.edu (C. Lin).

https://doi.org/10.1016/j.jesp.2023.104495
Received 5 February 2022; Received in revised form 12 October 2022; Accepted 28 May 2023

both traits and mental states are not directly observable (DeYoung et al., 2010; Haynes & Rees, 2006), much prior literature has examined how people infer these latent social features independently from observable features, such as momentary facial expressions and trait-implying behavior (Korman & Malle, 2016; Mende-Siedlecki, 2018; Uleman et al., 2008; Young & Saxe, 2009). However, when both psychological processes are studied in the same context, observable features may not be the only way that people make these inferences. Here, we investigate the possibility that people predict others' situation-specific mental states in a wide range of contexts using information about others' hidden traits, and infer others' enduring traits using information about others' momentary mental states.

2. Inferences of enduring traits

Research over past decades has shown that people draw inferences of others' traits from trait-implying behavior and do so without intentionally thinking about it (Hackel et al., 2020; Uleman, Newman, & Moskowitz, 1996; Winter & Uleman, 1984). For instance, people infer a person to be clumsy when observing the person bump into a chair. These trait inferences are shown to be modified by various factors. For example, given the same behavioral information, people are more likely to make trait inferences from the behavior when the target is psychologically distant, like a stranger (Rim, Uleman, & Trope, 2009; Trope & Liberman, 2010). People's own personality traits also influence how often they make trait judgments of others and how positive or negative those judgments tend to be (Tormala & Petty, 2001; Wood, Harms, & Vazire, 2010).

People also judge each other's traits from enduring cues, such as their facial structures. For example, people perceive baby-faced individuals to be warm, upon viewing the individuals' emotionally neural facial images (Oosterhof & Todorov, 2008; Zebrowitz, 2017). These trait judgments based solely on faces are shown to predict consequential real-world outcomes, such as hiring decisions, election results, and courtroom sentencing (Hamermesh, 2011; Jaeger, Todorov, Evans, & van Beest, 2020; Lin, Adolphs, & Alvarez, 2017, 2018; Rule & Ambady, 2011; A. Todorov, 2005; Wilson & Rule, 2015). However, trait judgments from faces are clearly not perfectly accurate. Debate remains regarding whether they hold kernel of truth (Bonnefon, Hopfensitz, & De Neys, 2015; Foo, Sutherland, Burton, Nakagawa, & Rhodes, 2021; Penton-Voak, Pound, Little, & Perrett, 2006), or whether they are purely reflections of perceivers' stereotypes and biases (Krosch, Berntsen, Amodio, Jost, & Van Bavel, 2013; Oh & Todorov, 2020; Xie, Flake, Stolier, Freeman, & Hehman, 2021).

3. Inferences of momentary mental states

As with inferring others' traits, much research has shown that people spontaneously infer others' mental states (Aviezer, Trope, & Todorov, 2012; Teufel, Fletcher, & Davis, 2010; Tomasello, Carpenter, Call, Behne, & Moll, 2005). For example, one might attribute that a person is stressed if they frown and scratch their head. The ability to infer others' emotions, goals, and beliefs emerge at an early age. For instance, beginning from 7 months of age, infants are able to categorize facial configurations that are associated with different emotions (Ruba & Pollak, 2020; Skerry & Spelke, 2014). At around one year of age, infants are able to infer goals from familiar actions as the action is being performed (Cannon & Woodward, 2012; Elsner & Adam, 2021). The ability to infer others' mental states is key to human social learning, such as learning whether to approach a situation and whether someone is a helper (Gweon, 2021; Hamlin, Wynn, & Bloom, 2007; Wu et al., 2021).

People are also able to infer others' mental states even without observing any dynamic cues or even without directly observing the target person at all. For instance, by simply looking at a static image of the target person's eye region, people are able to decode what the person is feeling or thinking about (e.g., concerned, playful) (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997; Lee & Anderson, 2017). By simply observing the contextual information, such as the surrounding scene (e.g., in a crowded bus) and/or the surrounding people (e.g., the interacting partners of the target), people are able to infer the target person's mental states (Barrett, Mesquita, & Gendron, 2011; Chen & Whitney, 2019; Martinez, 2019).

4. Interactions between inferences of traits and mental states

Long research traditions have examined inferences of traits and mental states separately. Nevertheless, the potential interactions between these two types of inferences could not go unnoticed. Inferences of traits based on neutral faces are found correlated with those faces' structural resemblance to emotional expressions (Said, Sebe, & Todorov, 2009). For instance, people judge others as more caring when the structure of others' neutral faces resemble happier expressions. However, it is unclear whether people make these trait inferences (e.g., caring) directly based on the static facial structure, or via the inferred mental states (e.g., happy) from the facial structure. Emotion judgments from faces are also influenced by the perceived traits of the faces. For instance, digitally manipulating a face to look less trustworthy causes people to perceive it as angrier (Oosterhof & Todorov, 2009).

Evidence of the interactions between inferences of traits and mental states is also found in contexts beyond face perception. In one neuroimaging study (Thornton, Weaverdyck, & Tamir, 2019), researchers obtained brain activity patterns when participants thought about different features of famous people such as their attitudes and preferences (person-patterns). They also obtained brain activity patterns when another group of participants thought about various mental states (state-patterns). They found that person-patterns could be reconstructed by summing the state-patterns, weighted by how frequently each famous person was thought to experience each mental state. This finding suggests that neural representations of others' enduring characteristics may be composed of representations of their habitual mental states.

5. The current research

The primary goal of the present investigation is to understand the nature of the causal interactions between mental state inferences and trait inferences. Before we examine these causal patterns, we first establish whether the correlations between traits and states are robust and generalizable. Building on prior findings that show associations between trait and mental state inferences in static faces (Oosterhof & Todorov, 2009; Said et al., 2009) and famous target people (Thornton, Weaverdyck, & Tamir, 2019), we test whether these associations appear in more naturalistic contexts, including movie viewing (Study 1) and judging personally familiar others (Study 2). Having shown that these associations are indeed robust in these more naturalistic contexts, we then examine the causal links behind these associations. To this end we test whether mental state frequencies causally influence trait inferences, and whether trait inferences influence situated mental state predictions. In the processes, we examine the specificity of these causal patterns to understand how individual dimensions of mental state and trait affect one another. The findings allow us to reunify two key domains of social cognition that have been artificially separated in much prior research.

Prior to testing causal hypotheses about trait-state interactions, it is essential to establish whether the correlations between these domains are generalizable. We consider three different aspects of generalizability here. First, the artificial stimuli used in prior research (e.g., isolated neutral faces) unnaturally constrain the social inferences people make and the information available to support those inferences. These constraints may eliminate, or conversely, manufacture the connection between trait and mental state inferences (Jolly & Chang, 2019; Schmuckler, 2001; Sonkusare, Breakspear, & Guo, 2019). Therefore, understanding the potential interactions between trait and mental state inferences in contexts that are more relevant to real life, such as making judgments from naturalistic videos, is essential to understanding whether this interaction truly exists.

Second, trait and mental state inferences are independently influenced by psychological distance (Trope & Liberman, 2010). The more familiar the targets are, the less likely that people interpret the targets' behavior in terms of their enduring traits (Idson & Mischel, 2001). Instead, people interpret familiar others' behavior in terms of their likely mental states, and attribute mental states to them with greater granularity (Thornton, Weaverdyck, Mildner, & Tamir, 2019). How people might link trait and mental state knowledge when they are making inferences about strangers or famous people – the targets examined in much prior researcher – may be different from when they are making inferences about their friends and family members. Therefore, understanding the interactions between trait and mental state inferences in socially proximal targets is necessary to establish generalizability (Yarkoni, 2022) and applicability to everyday life.

Third, people describe others using hundreds of different trait words and mental state words (Lin et al., 2021; Tamir, Thornton, Contreras, & Mitchell, 2016). Some mental state words are conceptually more closely linked to trait words (e.g., "anxious" could describe a person's trait and mental state), while others are more distinct (e.g., "awe" describes a mental state only). Examining how inferences of traits and mental states might interact for unrepresentative subsets of traits or mental states may distort the overall picture of their relationship (Jolly & Chang, 2019). Therefore, characterizing the interactions between these two processes for traits and mental states that represent the comprehensive dimensions of person perception is critical to generating unbiased conclusions.

After establishing that the associations between trait and state inferences are indeed robust and generalizable, we can safely proceed to our primary goal of testing the causal connections between them. Prior correlational findings between trait and mental state inferences are not sufficient to demonstrate causal links between these inferences. For instance, a correlation between trait and mental state inferences might arise because both types of inferences are correlated with a third confounding variable, such as a person's occupation. Therefore, establishing the causal effects between trait and mental state inferences is crucial to understanding whether and how people might use information about others' traits to infer others' mental states, and use information about others' mental states to infer their traits.

If there is direct causation between trait and mental state inferences, at least three potential causal patterns might obtain. First, it is possible that only mental state inferences causally affect trait inferences. That is, people might think that individuals who experience mental states in different situations in a specific way have a specific trait profile (i.e., trait profile is a function of mental state profile). Second, it is possible that only trait inferences causally affect mental state inferences. That is, people might think that individuals with a specific trait profile would experience mental states across different situations in a specific way (i. e., mental state profile is a function of trait profile). Third, it is possible that trait and mental state inferences causally influence each other bidirectionally. That is, people might think that individuals who experience mental states across different situations in a specific way would have a specific trait profile, and the reverse is also true.

Mathematically, the third causal pattern (bidirectional causation) is equivalent to the case where trait inferences are a function of mental state inferences, and conversely, mental state inferences are also a function of trait inferences. Formally, a variable y is regarded a function of variable x if given one value of x, we can predict exactly one value of y (Spivak, 2008). Following this definition, if y is a function of x, and that x is also a function of y, then x and y is a one-to-one correspondence, or formally, a strictly monotonic relation in which as one variable goes "up" the other must invariably do so as well (Clapham & Nicholson, 2014). This implies that if we observe bidirectional causation between trait and mental state inferences, then the relationship between them must be strictly monotonic. Strictly monotonic relationships can still be nonlinear (e.g., exponential) but knowing that the relationship between trait and mental state inferences is strictly monotonic would rule out a wide array of possible nonlinear relationships between them (e.g., quadratic, possible in the first and second causal patterns mentioned above). This would dramatically reduce the space of possible theories that future researchers would need to search to establish the exact form of the relationship between mental state and trait inferences.

To conclusively address these open questions, we conducted four pre-registered studies that combined a range of rigorous methods, including naturalistic paradigms, stimulus optimization techniques, statistical learning experiments, and representational similarity analysis. We first characterized the correlations between trait and mental state inferences in contexts that are more relevant to real life. In Study 1, participants formed impressions of unfamiliar targets in naturalistic videos. In Study 2, participants reported their impressions of the traits and mental state frequencies of familiar individuals in real life such as their friends and family members. In both studies, we examined how the inferences of traits and mental state frequencies of these targets were correlated. After establishing the correlations between trait and mental state inferences in more naturalistic contexts, we investigated whether these correlations were driven by causation. We conducted a pair of carefully controlled experiments. In Study 3, we manipulated the target people's perceived mental state frequencies via a statistical learning task, in which participants learned the mental states that the target experienced across a range of situations. We measured how different perceived mental state frequencies led to different trait attributions of the targets. In Study 4, we manipulated target people's perceived traits using biographies that contained facial images and text descriptions. We measured how different perceived traits led to different mental state attributions across a range of situations. To maximize generalizability, the traits, mental states, and situations used in all studies were representatively sampled. All studies were pre-registered on Open Science Framework. We report all measures, manipulations, and exclusions in these studies.

6. Study 1

We first investigated whether trait and mental state inferences might be correlated in a relatively naturalistic paradigm. Participants in this study viewed Hollywood movies and home videos and formed impressions of targets. This study imposed less constraints on what types of social inferences people made, and the information available to make those inferences. The videos used in the present study are more naturalistic compared to prior research in the following three aspects. First, these videos portray people dynamically. These dynamic expressions and movements provide additional information (beyond static features), which people may use to make trait and mental state inferences in naturalistic contexts. Second, these videos present various information streams simultaneously, including one's face, body, movements, interactions, and situations. These multi-modal features resemble the complexity people face when making social inferences in the real world. Third, compared to text descriptions of behaviors or situations, these videos depict information more realistically, such as describing a busy background with recording of an actual busy street. These realistic presentations allow participants to make social inferences in a similar way as they do in everyday life. Therefore, Study 1 will reveal whether trait and mental state inferences are correlated in contexts that are more relevant to real life (Jolly & Chang, 2019; Schmuckler, 2001; Sonkusare et al., 2019).

6.1. Method

The pre-registration of all methods for Study 1 can be accessed on the Open Science Framework: https://osf.io/5xpng?view_only=6a0af0e751554948897cf150c139a34b.

6.1.1. Participants

Participants were recruited from the online platform, Prolific.co. Participants were required to be aged 18 and older, with normal or corrected-to-normal vision, at least high school education, a good performance history (approval rate ≥ 98% on Prolific), and English fluency. All participants provided informed consent in a manner approved by the Committee for the Protection of Human Subjects of the authors' affiliated institution.

We determined the sample size based on formal power analyses (targeting $\alpha = 0.05$ and 95% power) and participant exclusion contingencies (Table 1). In Study 1, we based our power analysis on our main hypothesis that mental state frequencies would be associated with post-video trait ratings across individual participants (see Procedures below). We targeted a medium effect ($d = 0.36$), as defined by a recent survey of the actual distribution of effect sizes in psychology research (Lovakov & Agadullina, 2017). A power analysis using one-sample two-sided *t*-test indicated that we would need at least 103 participants. Assuming an equal number of participants across experiment modules (see Materials below) and an exclusion rate of around 15%, we determined to recruit 124 participants in total. We also planned that, if the actual exclusion rate turned out to be over 15%, we would recruit participants until the final sample size after exclusion (see criteria below) reached 124 participants.

In Study 1, we processed the data according to the following criteria. Participants were excluded if they failed >3 attention checks, or gave the same rating to over 90% of the trials. Responses for a particular target from a given participant were excluded if the participant failed the attention check for that target, or had more than half of the ratings for that target with response times shorter than 500 milliseconds. Besides these preregistered exclusion criteria, due to unforeseen data quality and technical issues (e.g., some participants submitted their tasks without completing, some participants had difficulty loading the videos), we supplemented additional exclusion criteria after data collection began but the data had not been analyzed. Specifically, we excluded (i) participants who could not have properly completed the study (whose submission time was shorter than the total video playback time); (ii) participants who took longer than 1 h to complete the study (an indication of internet or video playback issue); and (iii) any trait rating with a response time shorter than 500 ms. According to these criteria, $n = 27$ participants were excluded, resulting in the final sample size of 124 participants (53 women, 70 men, 1 non-binary; Age [M = 27, SD = 10]) as planned in our pre-registration (Table 1).

We performed a sensitivity power analysis for our main hypothesis that trait and mental state inferences would be correlated based on individual-level data (see Statistical Analysis below). The sensitivity analysis showed that with our final sample size of 124 participants, the minimum effect size that this study could detect with 80% power and $\alpha = 0.05$ would be $d = 0.25$ (and $r = 0.12$ in Pearson's correlation) based on a one-sample two-sided *t*-test.

6.1.2. Materials

Study 1 used 51 videos from a previously published study (Chen & Whitney, 2019). All videos were colored, and muted. These videos varied in length (ranging from 28.4 s to 179.2 s, with a median length of 86.3 s), genres (Hollywood movies, documentaries, home videos), and the number of characters involved (one, two, or more characters). The database originally included 60 videos; 9 were excluded from our study for contents involving violence, (implications of) bodily harm, and intense stress.

Compared to stimuli used in prior research, these videos (i) present target people dynamically and (ii) incorporate complex contextual information (iii) in a realistic way that resembles how people observe each other in everyday life. These three naturalistic aspects critically expand what information participants could freely observe or infer about the targets. For instance, participants could not only infer the targets' traits or mental states – variables that us researchers intend to study, but also obtain information about the targets' occupations, social relations, and so on as in real life. These naturalistic aspects also critically expand how participants may freely infer the targets' traits and mental states. For instance, participants could make these inferences based not only on static faces, but also on clothing, body poses, and so on as in real life.

Since it would take a long time for a participant to view and rate all 51 videos, these 51 videos were randomly grouped into four separate modules with the constraint that clips extracted from the same movie were not assigned to the same module (13 videos in Modules 1–3, and 12 videos in Module 4). We measured trait inferences using four traits (warmth, competence, femininity, youth), which were the core trait dimensions in person perception (Fiske et al., 2007; Lin et al., 2021). To mitigate the heterogeneity in how different participants may interpret a trait word, a one-sentence definition of each trait word was provided to participants. We obtained mental state inferences from the previous study (Chen & Whitney, 2019), where their participants provided continuous ratings of the targets on valence and arousal during the videos. A summary of the materials used in Study 1 is provided in the Supplemental Material (Table S1).

6.1.3. Procedures

Participants were randomly assigned to one of four modules. Each module contained a subset of the 51 videos that the participants watched as mentioned in the Materials section above. Participants were told that this was a study about inferring others' traits from videos. To understand how participants may update trait inferences as they gather more information about the target people's mental states, we measured trait inferences both before and after participants view the videos. Specifically, for each video (Fig. 1A), participants first viewed a still frame from the beginning of the video in which the target person was shown. The target person was highlighted in a red circle, and participants were instructed to pay attention to the target throughout the video. After participants had enough time to look at the target (5 s), the target was occluded. Participants provided trait ratings about the target based on their first impressions formed while the target was visible. Participants rated the target on four traits: warmth, competence, femininity, and youth (in a randomized order), using a 7-point Likert scale, anchored at 1 = not at all and 7 = very much. Next, participants viewed the video. After viewing the video, participants rated the target person on the four traits again (in randomized order). Finally, participants completed an

Table 1
Summary of sample size and demographics.

Studies	Power Analysis	Pre-registered Recruitment Plan	Total Recruited	Actual Excluded	Final Size	Gender	Age
Study 1	N ≥ 103	124 if exclusion <15%; otherwise, recruit until final sample size = 124	151	18%	124	53 women 70 men	M = 27 SD = 10
Study 2	N ≥ 156	164 if exclusion <5%; otherwise, recruit until final sample size = 156	183	15%	156	64 women 92 men	M = 41 SD = 11
Study 3	N ≥ 219	256 if exclusion <15%; otherwise, recruit until final sample size = 224	311	28%	224	139 women 85 men	M = 39 SD = 11
Study 4	N ≥ 219	256 if exclusion <15%; otherwise, recruit until final sample size = 256	433	40%	258	115 women 141 men	M = 39 SD = 11

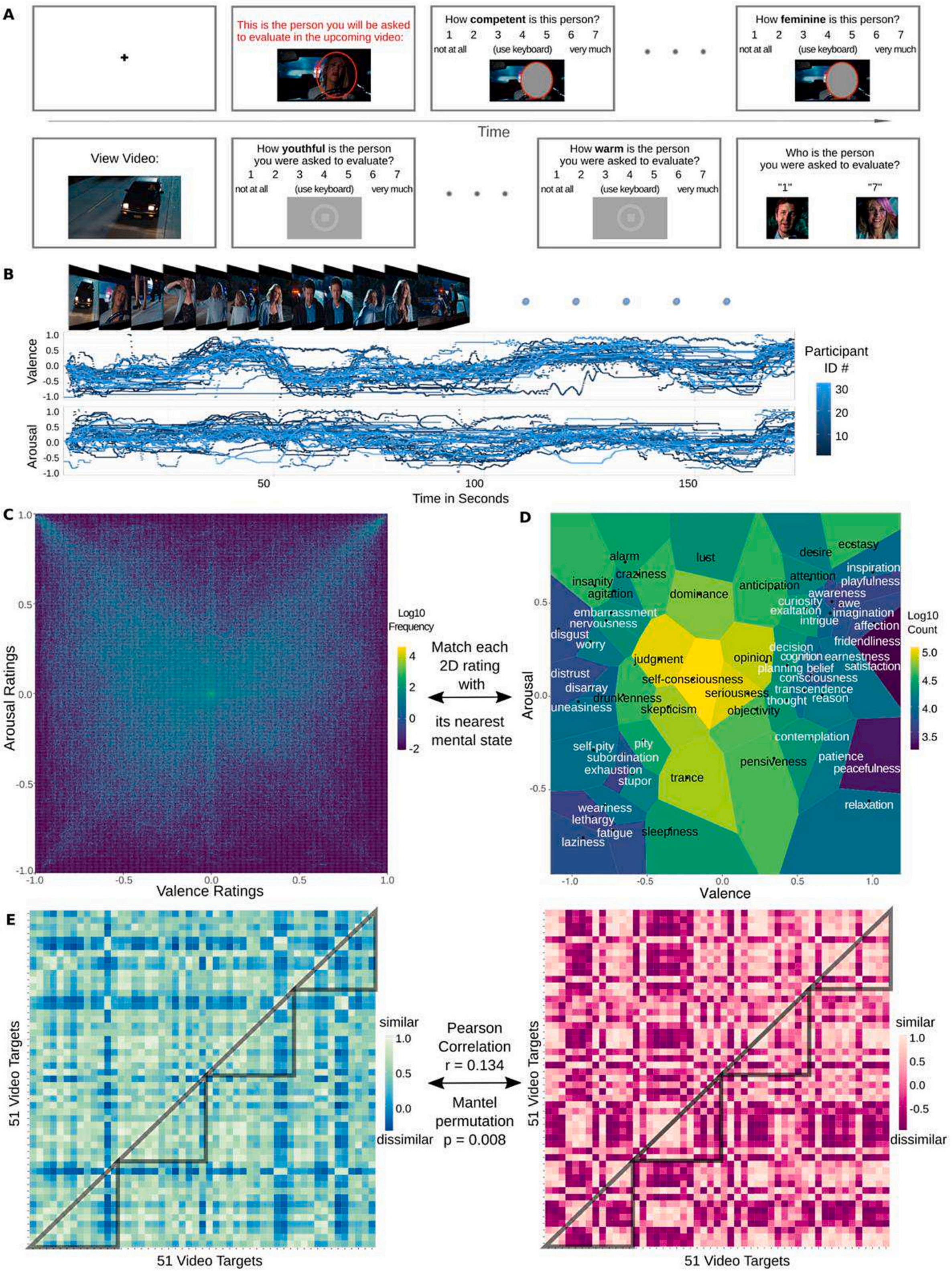


Fig. 1. Inferring traits and mental states from naturalistic videos. (*A*) Participants gave trait ratings to the target person before and after viewing the video. **(*B*)** An example of a video stimulus and the corresponding valence and arousal ratings (binned to every 100 ms) of the target person by the participants from a previous study (Chen & Whitney, 2019), who were different from those in (*A*). **(*C*)** Frequency of different valence and arousal ratings across all time points, all 51 targets in the videos, and all participants in (*B*). **(*D*)** Ratings in (*C*) were binned based on the nearest discrete mental states (black dots and text labels) in the valence-arousal space. **(*E*)** The Pearson correlations between targets calculated based on the frequency of matched mental states in (*D*) (left) and post-video trait ratings in (*A*) (right). This Pearson correlation was computed using the left and right heatmaps' unique pairwise correlations within the gray triangle areas ($N = 300$ pairs), which indicate four subsets of targets that were rated by four different groups of participants in (*A*).

attention check, in which they viewed images of two people and selected which one was the target person they were asked to evaluate.

6.1.4. Statistical analysis

We measured trait inferences of the targets in the videos from our participants' ratings directly (Fig. 1A). For mental state inferences, we used the valence and arousal ratings of the targets (Fig. 1B) collected in a previous study (Chen & Whitney, 2019). While valence and arousal are key dimensions for summarizing mental states (Tamir et al., 2016), people tend to use discrete mental state words (e.g., happy, planning) instead of valence or arousal to describe others' thoughts and feelings in everyday life. Therefore, we aimed to measure how frequently each target person was thought to experience a range of meaningful, named, discrete mental states. We did so by converting each valence-arousal rating of the target to the occurrence of a discrete mental state. Specifically, we first gathered a comprehensive set of 60 meaningful, named, discrete mental states from a prior study (Tamir et al., 2016). Then, we matched each momentary valence-arousal rating of the target to the mental state that shared the most similar valence-arousal rating among the 60 options (Fig. 1C-D). We then computed, for each target, the frequency of the 60 mental states matched to the target across the entire video and across all participants. We used these mental state frequencies for subsequent analyses.

To assess the correlation between inferences of traits and mental state frequencies, we used representational similarity analysis (RSA) (Kriegeskorte, Mur, & Bandettini, 2008). RSA helps bridge divides between different types of measures collected about the same items (e.g., brain-activity patterns, behavioral measures, computationally-extracted features about the same set of participants). It does so by quantitatively relating how similar the items (e.g., participants) are to each other in terms of each type of measures. Specifically, in our case, RSA was performed by first computing two similarity matrices (Pearson's correlation) between all pairs of targets. One similarity matrix between all pairs of targets was computed based on the targets' four trait ratings. The other similarity matrix between all pairs of targets was computed based on the targets' 60 mental state frequencies. Then, we computed the representational similarity (RS; Pearson correlation) between the trait-based similarities (Fisher's z transformed) and state-based similarities (Fisher's z transformed) across all unique pairs of targets in the same experiment module (Fig. 1E, triangle areas, n = 300 pairs). Pairs of targets belonging to two different experiment modules were not used for computing the RS. We deviated from our preregistration in this case because different groups of participants rated targets in different experiment modules, and therefore including pairs of targets from different modules would introduce perceiver-level variance that we did not intend to analyze.

We performed RSA using both aggregated trait ratings (to reduce noise in the data) and individual-level trait ratings (to avoid potential artifacts from aggregating the data). Mental state frequencies were derived only based on aggregated data from the previous study (Chen & Whitney, 2019). We used the Mantel test to assess the significance of the RS computed using aggregated trait ratings (Fig. 1E). The Mantel test is a permutation test that takes into account the dependencies among the items in a similarity matrix. For the RSs computed using individual-level trait ratings (Fisher's z transformed), its significance was assessed using one-sample two-sided t-test across all 124 participants. A significant RS would indicate that perceived traits and mental state frequencies are correlated when evaluating people in naturalistic videos.

To address the concern about converting the valence-arousal ratings to discrete mental states, we performed an exploratory analysis beyond preregistration. Prior research suggests that a valence-arousal value may not uniquely correspond to a mental state (Barrett, Mesquita, Ochsner, & Gross, 2007). Therefore, in this analysis, we did not convert the valence-arousal ratings to discrete mental states. Instead, we used these valence-arousal ratings to compute valence-arousal frequencies directly (Fig. 1C). Valence-arousal frequencies per target were quantified as the 2D kernel density based on the valence-arousal ratings. We computed the 2D kernel density using R function *kde2d*, with a sufficiently large number of grids, n = 10,000, across the space. After obtaining the valence-arousal frequencies for each target, we compared them with the trait ratings of the targets using RSA as before. This approach represented mental state frequencies along the two core mental state dimensions (valence and arousal) instead of using discrete mental states. It relied on the valence-arousal ratings directly measured from participants, avoiding the potential error introduced by converting these ratings to discrete mental states.

We performed four other preregistered analyses, whose methods and results are detailed in the Supplemental Materials. In brief, we analyzed (i) how strongly mental state frequencies were associated with post- versus pre-video trait ratings using RSA, (ii) how strongly and uniquely certain patterns of mental state frequencies predicted inferences of each single trait using LASSO regression with cross-validation, (iii) the consensus of trait inferences across participants using intraclass correlation coefficients, and (iv) the temporal stability between pre- and post-video trait inferences using Pearson's correlations.

6.2. Results

We observed a statistically significant correlation between post-video trait similarities and mental state frequency similarities across targets using RSA, when both measures were averaged across participants ($r = 0.13$, Mantel permutation $p = 0.008$; Fig. 1E). This correlation was also significant when the trait similarities were computed based on individual participants' trait ratings (mean $r = 0.12$, $t = 10.844$, $df = 123$, $p = 1.227 \times 10^{-19}$). These results suggest that people make associated trait and mental state inferences when judging unfamiliar targets in naturalistic videos.

We also found that, even without converting the valence-arousal ratings to discrete mental states, post-video trait similarities were significantly correlated with the valence-arousal frequency similarities ($r = 0.15$, Mantel permutation $p = 0.004$ for aggregated trait similarities; mean $r = 0.13$, $t = 11.346$, $df = 123$, $p = 7.420 \times 10^{-21}$ for individual participants' trait similarities). These results suggest that the observed frequency of mental states along two core mental state dimensions (valence, arousal) are linked to trait inferences. The correlations between trait inferences and mental state frequencies that were computed using abstract mental state dimensions versus discrete mental states were of a similar effect size. This finding suggests that frequencies of discrete mental states are an efficient and interpretable way for compressing time series of core mental state dimensions.

We found that mental state frequencies were not significantly more correlated with post-video trait ratings than pre-video trait ratings (mean $\Delta r = 0.01$, $t = 0.532$, $df = 123$, $p = 0.298$; see Supplemental Material). This finding suggests a directional relation: trait inferences of the targets from “thin slices” of observation at the beginning of the video (i.e., the still frame image of the target) might influence mental state inferences of the targets throughout the video. However, observing the targets' actual mental states during the video does not seem to update trait inferences accordingly. This finding highlights the question about the directional causation between inferences of traits and mental states, which we investigate further in Studies 3 and 4.

6.3. Discussion

Study 1 demonstrates that the correlation between inferences of traits and mental state frequencies exists in more naturalistic paradigms. The stimuli used here mimic many real-life encounters with strangers, in which people can observe others in complex, dynamic surroundings, and typically these observations are within a single situation and last for a short period of time. These findings are consistent with our hypothesis that people make sense of others' mental states using information about their traits.

7. Study 2

Study 1 indicated that inferences of traits and mental states may be connected in naturalistic contexts. However, the targets in that study were strangers, and many of the people we interact with in everyday life are personally familiar. Study 2 investigated whether the correlation between trait and mental state inferences generalizes to consequential familiar targets. It is unclear whether being familiar with the targets would encourage or discourage inferring their hidden traits from their hidden mental states, and vice versa (Fiske & Cox, 1979; Rim et al., 2009). For instance, it is plausible that people trust their assessments of familiar others' traits to be more accurate, which encourages using this trait information to make mental state inferences. However, it is also possible that people have additional information that is more relevant for inferring momentary mental states of familiar others, such as their preferences for specific subjects and situations, which discourages relying on general trait knowledge to make mental state inferences. To address this open question, Study 2 investigates the correlation between trait and mental state inferences when participants think about familiar individuals in real life such as their friends and family members.

7.1. Method

The pre-registration of all methods for Study 2 can be accessed on the Open Science Framework: https://osf.io/6yudg/?view_only=4bcca87690fd4cd2b81fb0b05327cbb9. There was no deviation from this pre-registration.

7.1.1. Participants

Participants were recruited from Amazon Mechanical Turk (MTurk) via Cloud Research (formerly known as TurkPrime) (Litman, Robinson, & Abberbock, 2017). Participants were required to be aged 18 and older, located in the US, with normal or corrected-to-normal vision, at least high school education, and a good performance history (approval rate $\geq$ 99% and submissions $\geq$ 50 on MTurk). All participants provided informed consent in a manner approved by the Committee for the Protection of Human Subjects of the authors' affiliated institution.

We based our power analysis on the hypothesis that the correlation between the overall similarity (see Procedures below) and mental state similarity across familiar people would be different from that between the overall similarity and trait similarity across familiar people. We targeted an effect size of 0.29, estimated from a previous dataset where 60 famous people were rated on 13 traits, 15 mental states, and overall similarity (Thornton, Weaverdyck, & Tamir, 2019). A power analysis using one-sample two-sided paired *t*-test indicated that we would need at least 156 participants. Assuming an exclusion rate of 5%, we determined to recruit 164 participants in total. We also planned that, if the actual exclusion rate turned out to be over 5%, we would recruit participants until the final sample size after exclusion (see criteria below) reached 156 participants, which was the minimum sample size indicated by the power analysis (Table 1).

In Study 2, we processed the data according to the following criteria. Participants completed different blocks of ratings: each similarity block included ratings of the overall similarity for 9 pairs of familiar people; each trait block included ratings of a specific trait for 10 familiar people; and each mental state block included ratings of a specific mental state for 10 familiar people. A block was excluded if all trials in the block had the same rating, or the response time for the block was too short (below 7200 milliseconds for the similarity block, 8000 milliseconds for the trait or mental state block). Participants were excluded if any of their similarity blocks was excluded, or >15% of their trait or mental state blocks were excluded. According to these criteria, $n = 27$ participants were excluded, resulting in the final sample size of 156 participants as planned (64 women, 92 men; Age [M = 41, SD = 11]) (Table 1).

We performed a sensitivity power analysis for our main hypothesis that trait and mental state inferences would be correlated based on individual-level data (see Statistical Analysis below). The sensitivity analysis showed that with our final sample size of 156 participants, the minimum effect size that this study could detect with 80% power and $\alpha = 0.05$ would be $d = 0.23$ (and $r = 0.11$ in Pearson's correlation) based on a one-sample two-sided t-test.

7.1.2. Materials

Study 2 measured trait inferences using 12 traits (warm, competent, feminine, youthful, trustworthy, dominant, attractive, extraverted, agreeable, conscientious, neurotic, openness to experience). These 12 traits were selected to represent the comprehensive dimensions of person perception based on previous literature (Fiske et al., 2007; Lin et al., 2021; Oosterhof & Todorov, 2008; Saucier & Goldberg, 1996). Mental state frequencies were measured for 12 mental states (awe, planning, lethargic, suspicious, calm, contemplating, interested, dread, jealous, indecisive, friendly, gloomy). These 12 mental states were selected using the maximum variation sampling procedure along the three core mental state dimensions (valence, rationality, social impact) from 160 candidate mental states (Tamir et al., 2016). Six of the 166 mental states in the original study were excluded prior to present selection due to either their ambiguity (cognition, consciousness, feeling, emotion), overlap with a selected trait (dominance), and not appropriate in the context of evaluating familiar people, such as family (lust). To mitigate the heterogeneity in how different participants may interpret a trait word or a mental state word, a one-sentence definition of each trait and mental state term was provided to participants. A summary of the stimuli used in Study 2 is provided in the Supplemental Material (Table S1).

7.1.3. Procedures

Participants were told that this was a study about how people perceive familiar others. First, participants wrote down the names of 10 familiar people; these names were used as stimuli in the rest of the task, but were not recorded. Participants then rated how similar these familiar people were to each other in a random order. These overall similarity judgments were based on whatever criteria each participant thought about spontaneously when asked to compare familiar people. These criteria may or may not be related to traits or mental states. Participants provided ratings using a 7-point Likert scale, anchored at 1 = very dissimilar and 7 = very similar.

Subsequently, participants rated the familiar people's traits and mental states. Questions about traits and mental states were grouped into two separate blocks. The order of the two blocks was randomized across participants. In the trait block, participants rated how well 12 sentences described each familiar person. These 12 sentences corresponded to 12 traits, each followed by a one-sentence definition (e.g., "This is a warm person. That is, this person is kind and loving in general."). Participants provided responses using a 7-point Likert scale, anchored at 1 = strongly disagree and 7 = strongly agree. The order of the 12 traits was randomized across participants. In the mental state block, participants rated how often they recorded each familiar person experienced 12 states of the mind across different situations. The 12 states of the mind corresponded to 12 mental states, each followed by a one-sentence definition (e.g., "How often does this person feel awe? That is, how often does this person feel respect or wonder?"). Participants provided responses using a 7-point Likert scale, anchored at 1 = never and 7 = always. The order of the 12 mental states was randomized across participants.

7.1.4. Statistical analysis

We used RSA to understand whether participants' evaluation of their familiar people in terms of traits and in terms of mental state frequencies was correlated. RSA was performed by first computing two similarity matrices for each participant across the 10 familiar people (Fig. 2). One similarity matrix between the 10 familiar people was computed based on the familiar people's trait ratings given by the participant. The other similarity matrix between the 10 familiar people was computed based on

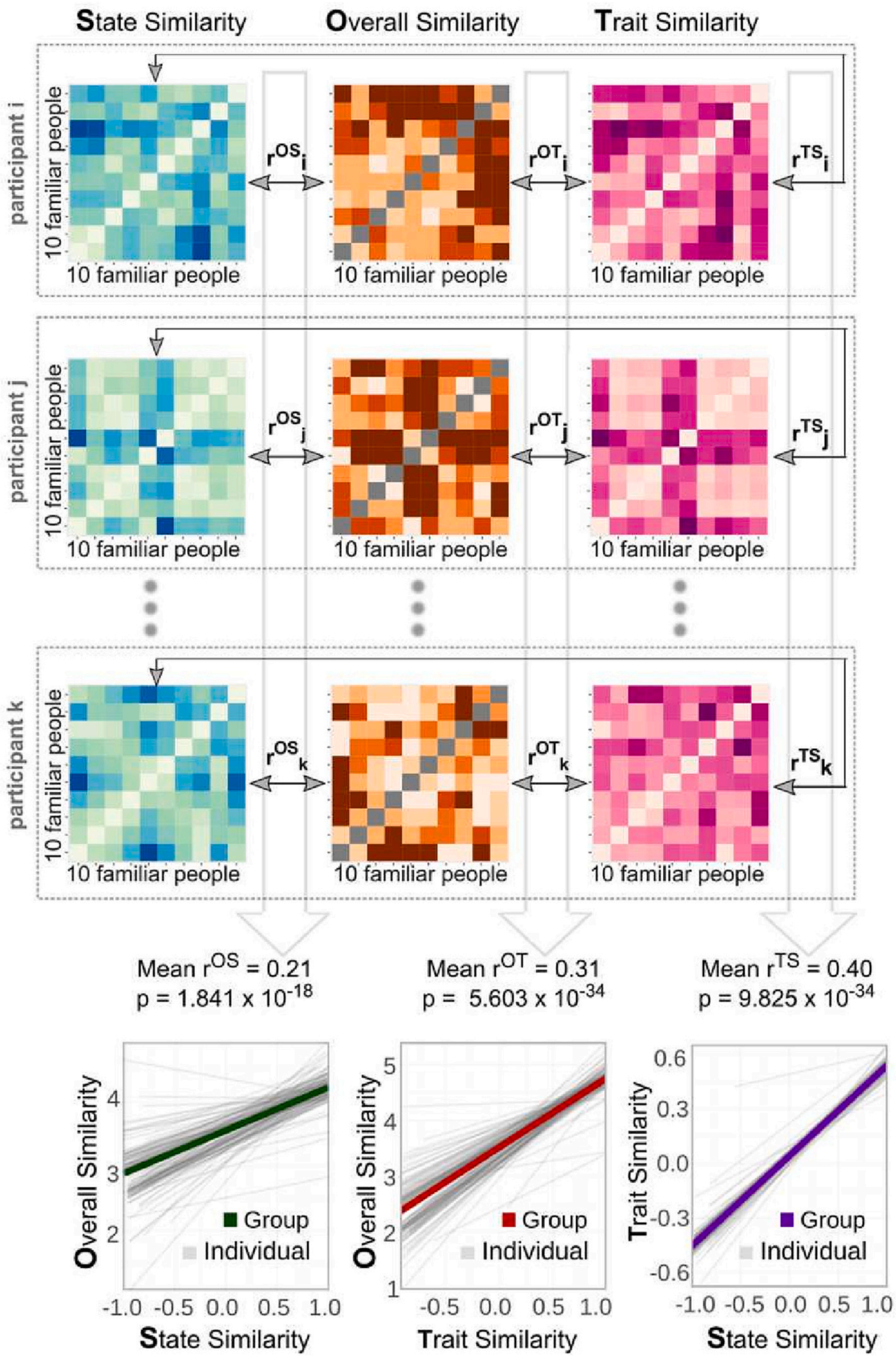


Fig. 2. Evaluating familiar others on traits and mental state frequencies. The similarities among 10 familiar people named by each participant were measured with overall similarity ratings (middle), as well as calculated based on their mental state frequency ratings (left) and their trait ratings (right). The Pearson correlations between the overall similarity and state similarity (r^{OS}, left line graph), overall similarity and trait similarity (r^{OT}, middle line graph), and trait similarity and state similarity (r^{TS}, right line graph) were computed for each individual participant (gray). One sided t-tests were performed for each of the three different correlations across participants (N = 156).

the familiar people's mental state frequencies rated by the participant. In both cases, we measured similarity using the Pearson correlation between familiar people – across traits, or across mental states. These trait similarities and state similarities were then linearized via Fisher's z-transform. Finally, we computed the RS (Pearson correlation, r^{TS}) between the trait similarities and state similarities across all unique pairs of the 10 familiar people (i.e. the lower triangle of similarity matrices) for each participant (Fig. 2, right line graph). The significance of these RSs across participants was assessed with one-sample one-sided *t*-test across participants (N = 156 participants). A significant result would indicate that the friends and family who are perceived to have similar traits are also thought to experience similar mental states with similar frequencies.

We also examined whether ratings of each trait across familiar people could be predicted by specific patterns of their mental state frequencies, and likewise, whether the perceived frequencies of each mental state across familiar people could be predicted by specific profiles of their traits. We answered this question using Ridge regression with cross-validation. The methods and results of this analysis are detailed in the Supplemental Materials.

To understand how relevant traits or mental states were to the spontaneous criteria that participants used for evaluating the overall similarity between familiar people, we performed two analyses: RSA (explained here) and variance partition analysis (see Supplemental Material). For RSA, we computed the Pearson correlation between the overall similarities and trait similarities (r^{OT}) across all unique pairs of familiar people per participant (Fig. 2, middle line graph). We also computed the Pearson correlation between the overall similarities and state similarities (r^{OS}) across all unique pairs of familiar people per participant (Fig. 2, left line graph). All of these correlations were z-transformed. We tested whether the RSs between overall and trait similarities (r^{OT}), and the RSs between overall and state similarities (r^{OS}) were significantly greater than zero across participants. The significance of each type of RSs was assessed with one-sample one-sided *t*-tests across participants ($N = 156$ participants). The significance of the difference between the two types of RSs was assessed with two-sided paired t-test across participants.

7.2. Results

Using RSA, we found that the more similar people were perceived to be in terms of traits, the more similar they were perceived to be in terms of mental state frequency ($r = 0.40$, $t = 15.872$, $df = 144$, $p = 9.825 \times 10^{-34}$; Fig. 2). These results indicate that when people evaluate familiar others in real life, they link the information about their traits and the information about their mental states.

We also found that the overall similarities between familiar people were associated with both trait similarity ($r = 0.31$, $t = 15.749$, $df = 151$, $p = 5.603 \times 10^{-34}$) and state similarity ($r = 0.21$, $t = 9.944$, $df = 148$, $p = 1.841 \times 10^{-18}$). This association was stronger between overall similarity and trait similarity than between overall similarity and state similarity (mean $\Delta r = 0.10$, $t = 5.817$, $df = 144$, $p = 3.727 \times 10^{-8}$ across participants). These findings were also confirmed by the variance partition analysis (see Supplemental Materials). These results indicate that, when thinking about how similar friends and families are, people spontaneously use knowledge about their traits and mental states, with trait knowledge potentially playing a larger role than mental state knowledge.

7.3. Discussion

Study 2 demonstrates that the correlation between inferences of traits and mental state frequencies exists when people evaluate personally relevant individuals in real life. These findings critically extend prior research: even when people have rich information about familiar targets for inferring their traits and mental states respectively, their evaluations of traits and mental states are still reliably associated. The correlation we observed in Study 2 ($r = 0.40$) was considerably larger than that in Study 1 ($r = 0.13$), suggesting that people connect these two types of hidden information more closely as they learn more about the targets (Smith et al., 2006; Thornton, Weaverdyck, Mildner, & Tamir, 2019). However, there are other differences between the two studies that might explain this effect size difference, such as Study 1 using a between-subject design and Study 2 using a within-subject design.

8. Study 3

Study 1 and Study 2 together show that trait and mental state inferences are reliably correlated in naturalistic and practically relevant contexts. To understand whether and how people use trait and mental state knowledge to inform each other, the crucial next step is to test the causal relationship between them. Study 3 examined whether mental state knowledge causally influences trait inferences. We manipulated the frequency with which the target people experienced different mental states across a range of situations (without describing those targets' traits). Participants learned the mental states of these different targets in a statistical learning task. We then measured how the manipulation of mental state frequency shaped subsequent trait inferences.

8.1. Method

The pre-registration of all methods for Study 3 can be accessed on the Open Science Framework: https://osf.io/p9m7a/?view_only=3773f7bf07bc47a39a700f735d958fbf. There was no deviation from this pre-registration.

8.1.1. Participants

Participants were recruited online from MTurk via Cloud Research (Litman et al., 2017). Participants were required to be aged 18 and older, located in the US, native English speakers, with normal or corrected-to-normal vision, at least high school education, and a good performance history (approval rate ≥ 99% and submissions ≥ 50 on MTurk). All participants provided informed consent in a manner approved by the Committee for the Protection of Human Subjects of the authors' affiliated institution.

We based our power analysis on the hypothesis that the similarity across targets computed from the manipulated variables (i.e., mental state frequencies) would be correlated with that computed from the measured variables (i.e., trait ratings) across all pairs of participants. We planned to test this hypothesis using correlation tests, where the number of observations was the number of unique participant pairs. However, these observations are not independent. For instance, how similar participant A's trait ratings are to those from participant B would not be independent of how similar participant A's trait ratings are to those from participant C. Therefore, we conservatively estimated that the number of independent observations in this correlation test to be its lower bound – the number of participants. We targeted a medium correlation ($r = 0.24$), as defined by a recent survey of the actual distribution of effect sizes in psychology research (Lovakov & Agadullina, 2017). A power analysis using the correlation test indicated that we would need at least 219 independent observations (i.e., 219 participants). Assuming an equal number of participants across experiment conditions (see Materials below) and an exclusion rate of around 15%, we planned to recruit 256 participants in total. We also planned that, if the actual exclusion rate turned out to be over 15%, we would recruit participants until the final sample size after exclusion (see criteria below) reached 224 participants (Table 1).

In Study 3, we processed the data according to the following criteria. Participants were excluded if i) their response times were shorter than 300 milliseconds in more than one third of the mental state learning trials, or more than three of the trait rating trials, ii) their accuracy rate in the third repeat of the mental state learning task was lower than that of the second repeat, or iii) their mental state frequency ratings in the manipulation check were not positively correlated with the actual manipulated mental state frequencies. According to these criteria, $n = 187$ participants were excluded, resulting in the final sample size of 224 participants (139 women, 85 men; Age [M = 39, SD = 11]), as planned in our pre-registration (Table 1).

We performed a sensitivity power analysis for our main hypothesis that making two target people appear to experience more similar mental state frequencies across situations would cause participants to attribute more similar traits to them. We planned to use representational similarity analyses to test this hypothesis (see Statistical Analysis below). Given our final sample size of 224 participants, the actual sample size in a representational analysis would be bounded by the nominal observation count (24,976 observations, the total number of unique pairs of similarity) and the minimal observation count (224 observations, the number of unique participants). With these upper and lower bounds of observations, the sensitivity analysis showed that the minimum effect size that this study could detect with 80% power and $\alpha = 0.05$ would be bounded by $r = [0.02, 0.19]$ based on a two-sided correlation test.

8.1.2. Materials

Study 3 manipulated participants' mental state attributions along the valence, rationality, and social impact dimensions. We focused on manipulating these three dimensions because they are the dimensions that summarize how people think about mental states (Tamir et al., 2016). Our manipulation generated target people (named either Mary or

James) in eight experiment conditions: all combinations of frequently or infrequently experienced positive, rational, and impactful mental states. We manipulated the mental state frequencies of the target people using a statistical learning task. Each trial in this task presented a situation and two mental states, and asked participants to choose which mental state better described how the target felt in the given situation. After participants made a choice, they were provided with feedback on how the target actually felt in the given situation, based on which participants learned the mental state frequency of the target.

The mental states and situations used in Study 3 were systematically selected as part of a separate, as-yet unpublished investigation (preregistration: https://tinyurl.com/2p4yvfhw). This investigation collected situation-state ratings for representative sets of 60 situations and 60 mental states. Each situation-state rating indicated how likely participants thought it would be for a person in that situation to experience that mental state (from 0% to 100%). These situation-state ratings were collected from 900 participants (290 female, 580 male, 30 declined to state; mean age = 26, range = 18–66) recruited from MTurk via Cloud Research (Litman et al., 2017). These situation-state ratings were averaged across participants to produce the values used for stimulus selection in our Studies 3 and 4. Specifically, in Study 3, for each situation we aimed to select two mental states to be the choice options that are i) both highly likely to occur in the situation for an average person (without considering our manipulation) and ii) most different from each other along the three mental state dimensions. To do so, for each situation, we first identified the five mental states that were rated as most likely to occur in the situation based on the situation-state dataset. From these five mental states, we then selected the two mental states that were most

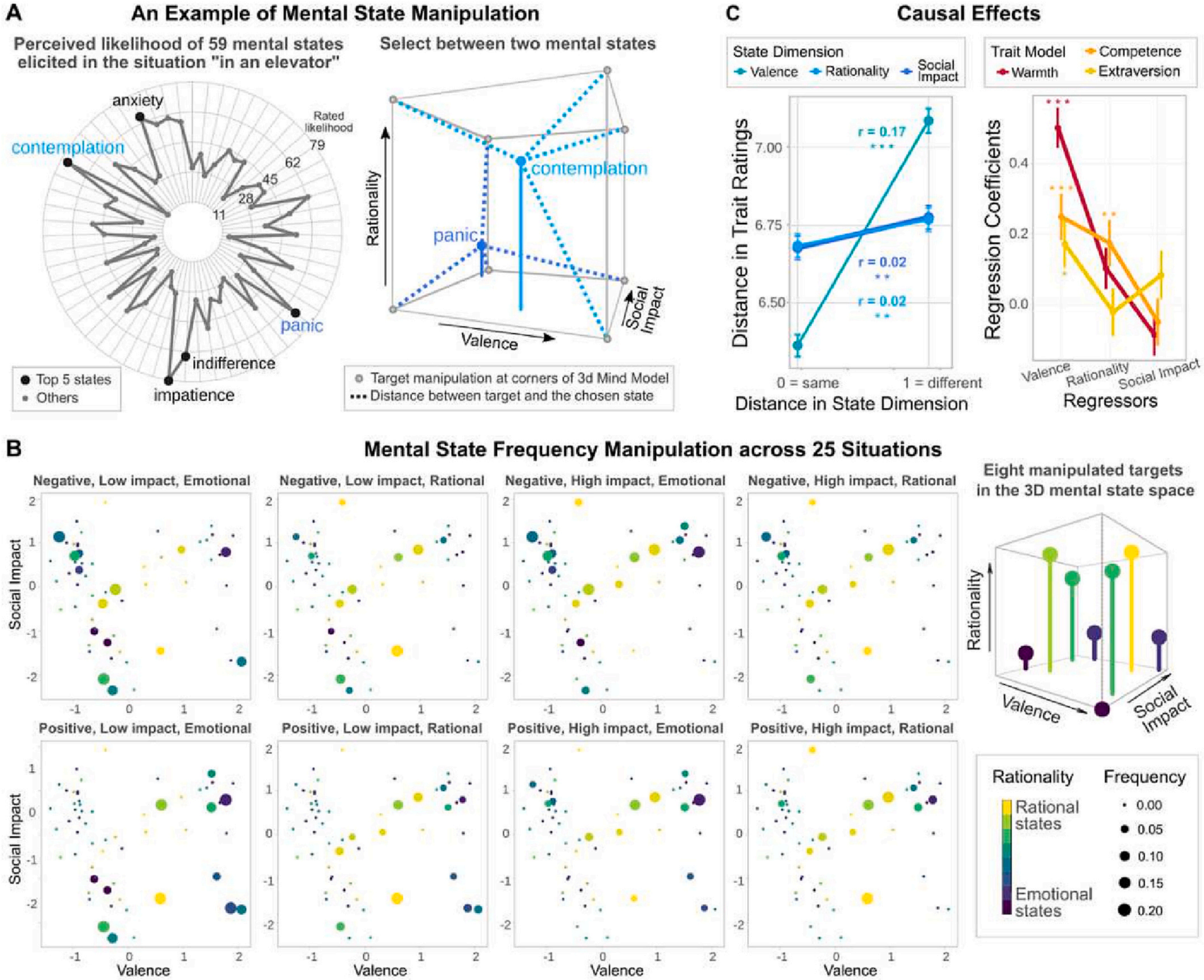


Fig. 3. Manipulations of mental state inferences change trait inferences (Study 3). (A) Left: Two choice options per situation. They were the top two mental states (colored labels) rated most likely in the given situation and furthest apart in the three-dimensional mental state space among a comprehensive set of 59 mental states. Right: Determination of the correct answer for a target's mental state in the given situation. It was the option closer to the extreme manipulation of the target (corners). (B) 2D Plots: The manipulated frequency of the target people with which they experienced 59 mental states (dots) across the optimally selected set of 25 situations (that maximize the dissimilarity among target people). The 3D plot: Distribution of the eight target people's mental state frequencies in the three-dimensional mental state space. (C) Left: The Pearson correlations between the Euclidean distance in mental state manipulation and the Euclidean distance in trait ratings across all pairs of participants (N = 24,976); error bars indicate 95% confidence intervals. Significance assessed with Mantel permutation tests. Right: Coefficients from linear regressions of trait dimension scores on the manipulation along three mental state dimensions (coordinates in the 3D plot of (B)) across all participants (N = 224). Error bars indicate one standard error of the coefficients. Asterisks indicate statistical significance: *** for $p < 0.001$, ** for $p < 0.01$, * for $p < 0.05$.

different along the valence, rationality, and social impact dimensions (Euclidean distance). The selected mental states become the two choice options in the learning trial for the corresponding situation (Fig. 3A, left).

After determining the two choice options per situation, we next determined which of the two options (mental states) each target person should be manipulated to feel (i.e., the correct answer). We aimed to select these correct answers to make the target people differ maximally from each other in their mental state experiences across situations (Fig. 3B, the 3D plot on the right). The targets would be most different if they were each most extreme in all three mental state dimensions – that is, all combinations of never or always experience positive, rational, and impactful mental states (Fig. 3B, the eight corners of the 3D plot). Therefore, we used these extreme cases as the reference for selecting the correct answers. Specifically, between the two choice options for a given situation, the correct answer for a target person in a given experiment condition (e.g., infrequently experiences positive, rational, and impactful mental states) was the option closer to the extreme of that experiment condition (e.g., never experiences positive, rational, and impactful mental states) (Fig. 3A, right).

After determining the correct mental state for each target in each situation, we computed the (correct) mental state frequency for each target across any subset of situations. We aimed to select a subset of situations that i) maximized the difference in mental state frequencies across targets, ii) and minimized the correlation among the three mental state dimensions. To do so, we targeted a subset with 20 to 50 situations. We deemed a subset size within this range to be appropriate based on two considerations: first, the number of situations should be big enough to be diverse and allowed for sufficient variance in the targets' mental state experiences; second, the number of situations should be small enough that the study remained within a reasonable length even when repeating all situations multiple times (for participants to learn about the target's mental state frequency). For each subset size, we computed the target people's mental state frequencies for all possible combinations of situations. Then, for each combination of situations, we computed i) the Euclidean distance between the mental state frequencies across all targets, and ii) the correlation between the targets' state-dimension scores. The state-dimension scores were the sum of the PC scores on that state-dimension across mental states (obtained from (Tamir et al., 2016)) weighted by how frequently the target was manipulated to experience those mental states. Results showed that the subset of 25 situations (Fig. 3B) was the optimal subset that satisfied our two selection aims.

After participants learned the manipulated mental states of the target people, Study 3 measured the inferences they made about the targets on 13 traits (agency, agreeableness, attractiveness, competence, conscientiousness, dominance, experience, extraversion, intelligence, neuroticism, openness, trustworthiness, warmth). These 13 traits were selected by a prior study to represent trait dimensions from four popular theories of person perception (Thornton & Mitchell, 2018). A summary of the stimuli used in Study 3 is provided in the Supplemental Material (Table S1).

8.1.3. Procedures

Participants were randomly assigned to one of the eight experiment conditions. In each condition, participants first saw the name of the person that they were going to learn about. Subsequently, they completed a mental state learning task. In each trial of this task, participants viewed a situation presented in a brief sentence (e.g., imagine James is at the gym) with the person's name in it. They then selected which of two mental state options (e.g., fatigue, enjoyment) the person would be more likely to experience in the given situation. After participants made a choice, they received feedback on the screen indicating whether their answer was correct or that the target person actually experienced the other mental state. To make sure that participants were aware of the correct answer on each trial, they had to press the key corresponding to the answer before continuing. There were 75 trials in total, with 25 unique situations (see Materials above). Each unique situation repeated 3 times for participants to learn about the target people's mental state experiences. Using participants' choices in the repetitions of the same situation, we could assess whether participants had learned the manipulated mental state of the target from the feedback (i.e., more accurate as the situation repeated). The order of all trials was randomized, subject to the constraint that the same situation could not appear twice in a row.

After completing the mental state learning task, participants rated the target person on 13 traits. A one-sentence definition of each trait was provided to participants. Participants rated the traits using a 7-point Likert, anchored at 1 = not at all and 7 = very much. Finally, participants completed manipulation checks, in which they rated how often the target person experienced each of 22 mental states. These 22 mental states were all the mental states that appeared as options over the course of the learning task. Participants rated mental state frequencies using a 7-point Likert scale, anchored at 1 = never and 7 = always.

8.1.4. Statistical analysis

We analyzed the causal link from mental state to trait inferences (Fig. 3C) using two methods: RSA and linear regression. We first used RSA to understand whether participants who were assigned more dissimilar targets in terms of their mental state experiences would make more dissimilar trait attributions about these targets. RSA was performed by first computing two similarity matrices (reverse-coded Euclidean distance) across all pairs of participants (N = 24,976 unique pairs). One similarity matrix between all pairs of participants was computed based on the mental state manipulation of the participants' assigned targets. We quantified the mental state manipulation of the targets using two types of metrics: i) the target's mental state frequency across the 22 mental states that ever appeared as options in the learning task, and ii) three binary variables indicating whether the target was manipulated to infrequently or frequently experience positive, rational, and impactful mental states. The other similarity matrix between all pairs of participants was computed based on the trait ratings given by each participant (a vector of 12 trait ratings). Then, we computed the RS (Pearson correlation) between the state-based similarities and trait-based similarities across all unique pairs of participants. We assessed the significance of the RS using the Mantel test.

We next performed linear regression analyses to understand whether the causal link from mental state to trait inferences was dimension-specific. That is, whether the manipulation along a certain mental state dimension only influenced the inferences of traits along a specific trait dimension. We focused on three trait dimensions: warmth, competence, and extraversion. We studied these three dimensions because prior research showed that the 13 traits used in our present study could be summarized by these three trait dimensions (Thornton & Mitchell, 2018). We expected that these three trait dimensions would be differentially responsive to state-valence, state-rationality, and state-social-impact manipulations. This expectation was based on prior research showing that trait warmth describes the valence of the global impressions of others across different situations (Fiske et al., 2007; Wojciszke, Bazinska, & Jaworski, 1998); trait competence describes one's capability of self-control and rationality across different situations (Fiske et al., 2007; Waytz, Gray, Epley, & Wegner, 2010); and trait extraversion is found to associate with larger social network and greater social influence (Feiler & Kleinbaum, 2015; Pollet, Roberts, & Dunbar, 2011).

To test these hypotheses, we fit three regression models, one predicting trait warmth, one predicting trait competence, and the other predicting trait extraversion. For example, for the trait warmth model, the dependent variable was the trait-dimension score on warmth. The trait-dimension score on warmth for each participant was the sum of the principal component loadings on warmth across the 13 traits (obtained from (Thornton & Mitchell, 2018)), weighted by the participant's ratings of those traits in our study. The trait-dimension score on the other two dimensions was computed in a similar way.

Each model regressed the trait-dimension scores on the targets' mental state manipulation along the valence, rationality and social-impact dimensions across participants (N = 224 participants). The manipulation along each state dimension was quantified with the state-dimension score. The state-dimension score on valence per participant was the sum of the PC scores on valence across 22 mental states, weighted by how frequently the participant's assigned target was manipulated to experience those mental states. The state-dimension score on the other two mental state dimensions was computed in a similar way. These state-dimension scores could be visualized as the coordinates of the target people in the three-dimensional mental state space (Fig. 3B, the 3D plot on the right).

8.2. Results

Manipulation checks confirmed that participants did learn the mental state frequencies of the target people in the mental state learning task as intended: participants provided mental state frequency ratings after the learning task that were significantly correlated with the targets' actual manipulated mental state frequencies (mean $r = 0.57$, $SD = 0.22$ across participants).

RSA showed that the more similar the targets' mental state frequencies were (along all three mental state dimensions), the more similar trait ratings they elicited ($r = 0.13$, Mantel permutation p = 0.001). Moreover, this effect held true for each of the individual mental state dimensions separately ($r = 0.17$, Mantel permutation $p = 0.001$ for the state-dimension valence; $r = 0.02$, Mantel permutation $p = 0.003$ for the state-dimension rationality; $r = 0.02$, Mantel permutation $p = 0.002$ for the state-dimension social impact; Fig. 3C, left). These findings indicate that manipulating the frequency of mental states causally shapes trait inferences.

Next, we analyzed whether this causal link was dimension-specific. In particular, we examined whether state valence specifically affects inferences of trait warmth, state rationality specifically affects inferences of trait competence, and state social-impact specifically affects inferences of trait extraversion. Linear regressions showed that trait warmth was predicted only by state valence ($b = 0.50$, $p = 2.753 \times 10^{-15}$), a better predictor than state rationality ($\Delta|b| = 0.40$, 95%CI [0.24, 0.55], bootstrap resampling) and state social-impact ($\Delta|b| = 0.41$, 95%CI [0.20, 0.60]). Trait competence was predicted by both state rationality ($b = 0.17$, $p = 0.009$) and state valence ($b = 0.25$, $p = 2.173 \times 10^{-4}$), but state rationality was not a significantly better predictor than state valence ($\Delta|b| = -0.08$, 95%CI [−0.25, 0.10]) or state social-impact ($\Delta|b| = 0.10$, 95%CI [−0.04, 0.24]). Trait extraversion was predicted only by state valence ($b = 0.17$, $p = 0.013$); state social-impact was not a better predictor than state valence ($\Delta|b| = -0.08$, 95%CI [−0.23, 0.08]) or state rationality ($\Delta|b| = 0.04$, 95%CI [−0.09, 0.18]). These results indicate that the causal link from mental state to trait inferences is only partially dimension-specific: inferences of trait warmth is selectively affected by state valence. Contrary to our prediction, inferences of trait competence was not selectively affected by state rationality, and inferences of trait extraversion was not selectively affected by state social-impact. However, descriptively (Fig. 3C, right; examined column-wise), state rationality did affect inferences of trait competence more than trait warmth or extraversion, and state social-impact did affect inferences of trait extraversion more than trait warmth or competence.

8.3. Discussion

Study 3 demonstrates that inferences of mental states across a range of situations causally influence trait attributions. These findings support our hypothesis that people use information about others' mental states to inform their traits. They do so not only when judging people from their faces as shown in prior research, but when learning about a diverse set of mental states across a comprehensively sampled set of situations.

Different from our expectation, we did not observe the predicted dimension-specific causal links for state rationality and state social-impact. We believe this stems from our failure to anticipate the much larger overall effect of state valence on trait inferences compared to state rationality or state social-impact (examining the left part of panel C in Fig. 3). This much larger main effect of state valence made it effectively impossible to observe the prediction advantages of state rationality or state social-impact for any particular trait dimension, because state valence ended up as the most predictive state dimension for all three trait dimensions. This may reflect the fact that valence is simply a more important dimension of mental state representation than the other two dimensions, a finding we have observed in past studies (Thornton, Mark, & Tamir, 2017; Thornton & Tamir, 2020).

9. Study 4

Study 3 established that mental state inferences can affect trait inferences. This causal path alone is sufficient to explain the correlation between trait and mental state inferences observed in Studies 1 and 2. However, it remains unclear whether the causation between trait and mental state inferences is bidirectional. To test this possibility, Study 4 examined the other possible causal direction: from trait to mental state inferences. We manipulated the traits that participants attributed to different targets (without describing those targets' mental states). Participants learned about the targets' traits from biographies that contained the targets' facial images and a short paragraph describing how others thought about the targets. We then measured how this trait manipulation shaped inferences about the target's mental states across different situations.

9.1. Method

The pre-registration of all methods for Study 4 can be accessed on the Open Science Framework: https://osf.io/w9jbs/?view_only=43c722dc0e2243debc5ba8ee053c2397. There was no deviation from this pre-registration.

9.1.1. Participants

Participants were recruited online from MTurk via Cloud Research (Litman et al., 2017). Participants were required to be aged 18 and older, located in the US, native English speakers, with normal or corrected-to-normal vision, at least high school education, and a good performance history (approval rate ≥ 99% and submissions ≥ 50 on MTurk). All participants provided informed consent in a manner approved by the Committee for the Protection of Human Subjects of the authors' affiliated institution.

We based our power analysis on the hypothesis that the similarity between pairs of participants' assigned targets computed based on the manipulated variables (i.e., traits) would be correlated with that computed based on the measured variables (i.e., inferred mental state frequencies). We performed the power analysis using the same procedure and targeting the same effect size ($r = 0.24$) as in Study 3. The power analysis indicated that we would need at least 219 participants. Assuming an exclusion rate of around 15% and an equal number of participants across experiment conditions (there were a total of 64 versions of manipulation materials, see Materials below), we determined to recruit 256 participants in total. We also planned that, if the actual exclusion rate turned out to be over 15%, we would recruit participants until the final sample size after exclusion (see criteria below) reached 256 participants (Table 1).

In Study 4, we processed the data according to the following criteria. A mental state inference trial was excluded if its response time was shorter than 500 milliseconds. Participants were excluded if they i) had >5 trials excluded, ii) spent <6 s on viewing the biography (see

Materials below), iii) selected the same mental state in fewer than 10 out of the 30 test-retest trials, iv) selected the same mental state in >90% of the unique situations, or v) failed any manipulation check. According to these criteria, $n = 175$ participants were excluded, resulting in the final sample size of 258 participants (115 women, 141 men, 1 non-binary; Age [M = 39, SD = 11]), reaching the planned sample size (Table 1).

We performed a sensitivity power analysis for our main hypothesis that making two target people appear to have more similar traits would cause participants to attribute more similar mental state frequencies to them across situations. We planned to use representational similarity analyses to test this hypothesis (see Statistical Analysis below). Given our final sample size of 258 participants, the actual sample size in a representational analysis would be bounded by the nominal observation count (33,153 observations, the total number of unique pairs of similarity) and the minimal observation count (258 observations, the number of unique participants). With these upper and lower bounds of observations, the sensitivity analysis showed that the minimum effect size that this study could detect with 80% power and $\alpha = 0.05$ would be bounded by $r = [0.02, 0.17]$ based on a two-sided correlation test.

9.1.2. Materials

Study 4 manipulated participants' trait inferences about the target people along the warmth and competence dimensions using different biographies. We focused on manipulating the warmth and competence dimensions because they are the core trait dimensions in person perception (Fiske et al., 2007). Each biography included a studio portrait of the target person's neutral face, name (Mary or James), age (29 years old), state of residence (CA or NY), and a text description (Fig. 4A). Only the face and the text description were used for manipulation; the other elements were included to add verisimilitude and increase variance across the biographies. We manipulated trait inferences of the targets using face images and text descriptions because these two types of stimuli were shown to reliably elicit trait inferences from perceivers (Oosterhof & Todorov, 2008; Uleman et al., 2008). We selected the face images from a previously published study (Lin et al., 2021), which collected face ratings from human subjects for 100 comprehensively sampled traits and showed that warmth and competence were the top two dimensions that summarized those trait judgments. Based on the face ratings from that study, we selected four neutral faces of each gender that were most extreme along the trait dimensions of warmth and competence. An example of the face image is shown in Fig. 4A; other face images can be accessed from the original face databases (see Supplemental Material). We used neutral faces that exhibit divergent degrees of trait impressions to maximize the effect of perceived traits (the variable we intend to manipulate) and minimize the potential effect of facial expressions on mental state inferences (the variable we intend to measure). The text description stated how people who knew the target person generally thought about the target person's traits along the warmth and competence dimensions. We randomized the order of the descriptor for the two trait dimensions. The face image and the text description in each biography provided congruent manipulation. For instance, to manipulate a target person to be of low warmth and competence, a face that was rated low in warmth and competence was paired with text description that depicted the target person as low in warmth and competence. These manipulations resulted in 64 distinct biographies (2 genders × 2 names × 2 states of residence x 2 trait description orders). These biographies manipulated the targets' traits in four experiment conditions: targets who were of (i) low warmth and low competence, (ii) high warmth and low competence, (iii) low warmth and high competence, and (iv) high warmth and high competence.

Study 4 measured the subsequent mental state inferences (after participants viewed the biographies) using 6 representatively sampled mental states (Fig. 4B) and 30 systematically sampled situations. In each situation, participants were asked which of the six mental states best described how the target person felt. We decided to use six mental states as choice options across situations based on two considerations: i) the number of options should be small enough that the task remains easy for participants of diverse cognitive backgrounds; and ii) the number of options should be big enough to capture the variance in the inferred mental state frequencies across target people. As in Study 3, we based the selection of these situations and mental states on the situation-state dataset (pre-registration: https://tinyurl.com/2p4yvfhw). Using these data, we representatively sampled 6 mental states (Fig. 4B) from 59 mental states – one of the 60 mental states in the original study (feeling) was excluded in our present selection for its ambiguity. We used the maximum variation sampling procedure to sample these 6 mental states. This procedure maximized the dissimilarity among the selected mental states along the three core mental state dimensions (valence, rationality, social impact) (Tamir et al., 2016). After determining the six selected mental states, we next selected a subset of situations across which the six mental states were equally likely to be experienced by an average person (without the effect of trait manipulation). To do so, we selected the five situations that were rated as most likely to co-occur with each of the six mental states based on the situation-state dataset (30 situations in total, see Fig. 4C). A summary of the stimuli used in Study 4 is provided in the Supplemental Material (Table S1).

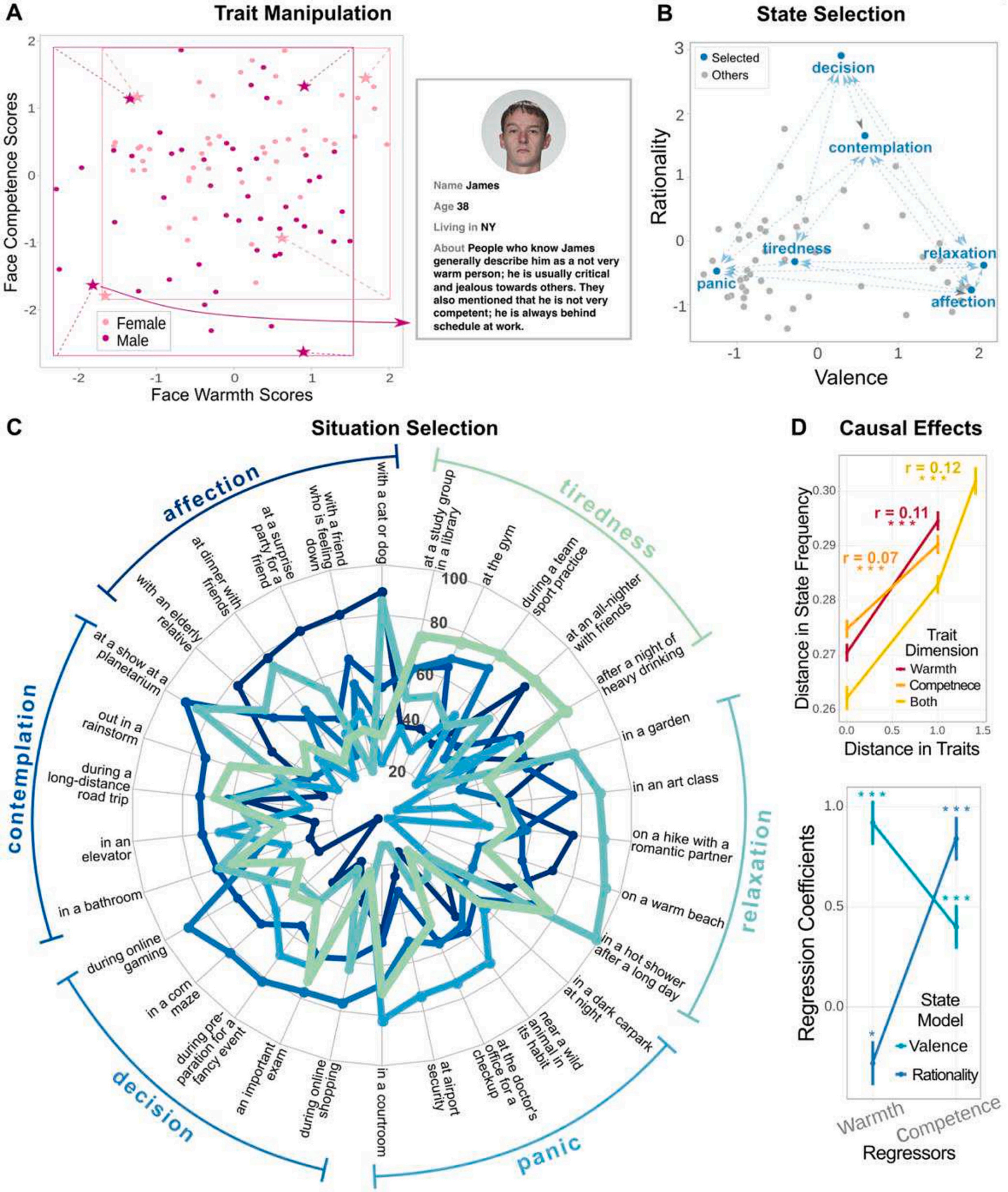
9.1.3. Procedures

Participants were randomly assigned to one of the four experiment conditions (see Materials above). In each condition, participants viewed a biography of the target person (Fig. 4A). After viewing the biography, participants viewed a series of situations one by one presented in brief sentences (e.g., imagine James is with a cat or dog), with the person's name and portrait remained on the screen. In each situation, participants selected which one of the six mental states best described what the target person felt. There were 60 trials in total, with 30 unique situations. Each situation was displayed twice for assessing the test-retest reliability of a participant's choice (used for excluding low-quality data, see Participants above). The 60 trials were randomized with the constraint that the same situation appeared at least one trial apart. After completing the mental state inference task, participants completed two manipulation checks. In these manipulation checks, participants rated the target person on warmth and competence, using a 7-point Likert scale, anchored at 1 = not at all and 7 = extremely.

9.1.4. Statistical analysis

We analyzed the causal link from trait to mental state inferences using two methods: RSA and linear regression. We first used RSA to understand whether participants who were assigned more dissimilar targets in terms of their manipulated traits would make more dissimilar mental state attributions across situations. RSA was performed by first computing two similarity matrices (via reverse-coded Euclidean distance) across all pairs of participants ($N = 33{,}153$ unique pairs). One similarity matrix between all pairs of participants was computed based on the trait manipulation of the participants' assigned targets. The trait manipulation of each target was a vector of 2 binary elements indicating whether the target was manipulated to be high or low on warmth and competence. The other similarity matrix between all pairs of participants was computed based on the mental state frequency with which the participants chose the six mental states across all trials. The mental state frequency per target per participant was a vector of 6 elements indicating the selection frequency of the six mental states by that participant for that target. Then, we computed the RS (Pearson correlation) between the trait-based similarities and state-based similarities across all unique pairs of participants. We assessed the significance of this RS using the Mantel test.

We next performed linear regression analyses to understand whether the causal link from trait to mental state inferences was dimension-specific. That is, whether the manipulation along a certain trait dimension influenced the inferred frequency of only some mental states along a specific mental state dimension. We focused on two mental state dimensions: valence and rationality. We expected that these two mental



Name James

Age 38

Living in NY

About People who know James generally describe him as a not very warm person; he is usually critical and jealous towards others. They also mentioned that he is not very competent; he is always behind schedule at work.

Fig. 4. Manipulations of trait inferences change mental state inferences (Study 4). (A) A biography includes two trait-manipulation-relevant information: a face image and a brief paragraph of description. Four faces from each gender with the most extreme scores on the two trait dimensions (corners) (Lin et al., 2021) were used for the manipulation. **(B)** Six mental states (plotted along two of the three mental state dimensions (Tamir et al., 2016)) were selected using maximal variation sampling procedure (maximizing summed distance between selections, dotted arrow lines) from a comprehensive set of 59 mental states. **(C)** A set of 30 situations (black text labels) were selected from a larger set as most likely to elicit one of the six selected mental states (5 for each mental state; scores indicated by the radial position of the dots). **(D)** Top: The Pearson correlations between the Euclidean distance in trait manipulations (dummy coded) and the Euclidean distance in mental state frequencies across all pairs of participants ($N = 33{,}153$); error bars indicate 95% confidence intervals. Significance assessed with Mantel permutation tests. Bottom: coefficients from linear regressions of mental state dimension scores on the manipulation along two trait dimensions across all participants (N = 258). Error bars indicate one standard error of the coefficients. Asterisks indicate statistical significance: *** for $p < 0.001$, ** for $p < 0.01$, * for $p < 0.05$.

state dimensions would be differentially responsive to trait-warmth and trait-competence manipulations. This expectation was based on prior research showing that trait warmth describes the valence of the global impressions of others across different situations (Fiske et al., 2007; Wojciszke et al., 1998), and that trait competence describes one's capability of self-control and rationality across different situations (Fiske et al., 2007; Waytz et al., 2010).

To test these hypotheses, we fit two regression models, one predicting state valence and the other predicting state rationality. For example, for the state valence model, the dependent variable was the state-dimension score on valence. The state-dimension score on valence for each participant was the sum of the principal component scores on valence across the six mental states (obtained from (Tamir et al., 2016)), weighted by the participant's selection frequency of those mental states in our study. The state-dimension score on rationality was computed in a similar way. Each model regressed the state-dimension scores on the targets' trait-warmth and trait-competence manipulations (each dummy coded: 0 for low, 1 for high) across participants ($N = 258$ participants).

9.2. Results

Manipulation checks confirmed that the biographies induced the intended trait attributions: the average warmth rating on a 7-point Likert scale was $M = 2.02$ ($SD = 0.68$) across participants assigned to low-warmth manipulation, and $M = 6.32$ ($SD = 0.66$) for high-warmth manipulation; the average competence rating was $M = 2.10$ ($SD = 0.75$) for low-competence manipulation, and $M = 6.27$ ($SD = 0.63$) for high-competence manipulation.

RSA showed that the more different the targets' manipulated traits were, the more different attributions of mental state frequency they elicited ($r = 0.11$, Mantel permutation $p = 0.001$ for the effect of manipulation along the trait-warmth dimension; $r = 0.07$, Mantel permutation $p = 0.001$ along the trait-competence dimension; $r = 0.12$, Mantel permutation $p = 0.001$ along both trait dimensions; Fig. 4D, top). These findings indicate that manipulating information about traits causally influences attributions of mental states across a range of situations.

Next, we analyzed whether this causal link was dimension-specific. In particular, we examined whether trait warmth specifically affects inferences of state valence, and trait competence specifically affects inferences of state rationality (see Statistical Analysis above). Linear regressions (Fig. 4D, bottom) indicated that state valence was predicted by both trait warmth ($b = 0.92, p = 1.185 \times 10^{-15}$) and trait competence ($b = 0.40, p = 2.613 \times 10^{-4}$), with trait warmth being a better predictor than trait competence ($\Delta|b| = 0.52$, 95%CI [0.25, 0.82], bootstrap resampling). State rationality was predicted by both trait competence ($b = 0.84$, $p = 9.109 \times 10^{-13}$) and trait warmth ($b = -0.28$, $p = 0.013$), with trait competence being a better predictor than trait warmth ($\Delta|b| = 0.56$, 95%CI [0.28, 0.84], bootstrap resampling; see SI). These results indicate that the causal link is indeed dimension-specific: trait warmth selectively affects inferences of mental states along the valence dimension, and that trait competence selectively affects inferences of mental states along the rationality dimensions.

9.3. Discussion

Study 4 demonstrates that inferences of traits causally shape inferences of mental states across a range of situations. These findings corroborate the hypothesis that people use information about others' traits to inform their mental states. People do so not only when judging emotions from faces as shown in prior research, but when thinking about a diverse set of mental states across a comprehensively sampled set of situations in our study.

10. General discussion

The present research investigated how people infer each other's enduring traits and momentary mental states. These two psychological processes have typically been examined independently. Here, we argue that removing this artificial separation is essential to better understand each process (Zaki, 2013). Given that trait and mental state inferences constantly co-occur in real life and utilize overlapping cues, we propose that there is a bidirectional connection between the two processes. Supporting this hypothesis, we observed reliable correlations between inferences of traits and mental state frequencies as participants formed impressions about targets from naturalistic videos (Fig. 1E) and evaluated personally familiar individuals (Fig. 2 and Fig. S1). These findings show that people naturally make associated trait and mental state inferences in naturalistic contexts. Critically, we also showed bidirectional causal effects between trait and mental state inferences using comprehensively sampled sets of traits, mental states, and situations (Table S1). Learning that two people experienced similar mental state frequencies across a range of situations caused participants to infer that they would have similar traits (Fig. 3C). Making two people appear to have similar traits caused participants to predict that they would experience similar mental states across a range of situations (Fig. 4D). These findings indicate that people not only use mental states and traits as independent tools for understanding others, but that people link traits and mental states together in a strictly monotonic way (e.g., more conscientious individuals are thought to experience the mental states of planning, contemplating, calm, etc. more frequently; see Fig. S1).

The sizes of the two causal effects were similar: $r = 0.12$ for the effect of mental state manipulation on trait inferences (Study 3), and $r = 0.13$ for the effect of trait manipulation on mental state inferences (Study 4). The correlation between trait and mental state inferences also showed a similar effect size: $r = 0.13$ when the targets are psychologically distal (Study 1). These effect sizes are in the 'medium' range for effects in preregistered between-subject designs (mean $r = 0.17$, $SD = 0.15$), according to recent empirical benchmarks (Schäfer & Schwarz, 2019). The effect sizes of the causal and correlational relationships between trait and mental state inferences may be subject to a natural ceiling due to the different levels of abstractions in trait and mental state inferences. For instance, perceivers may rapidly form an impression that a target is an intelligent person, but seldom infer that a target is in an intelligent mental state; instead, they are more likely to infer that the target is in a mental state of "planning" or "contemplating". That is, the higher level of abstraction in trait inferences in comparison to the lower level of abstraction in mental state inferences may have imposed an upper bound on how much variance in trait inferences could be explained by mental state inferences, and vice versa.

It remains an open question whether the relationships between trait and mental state inferences in real-world contexts would be larger or smaller than those found here. On the one hand, the manipulations we used in Studies 3 and 4 were small compared to all the information that could shift people's impressions of a target person's traits or mental state frequencies in real life. For instance, people's trait impressions about a target may change in real life not only from how the target's face looks and how others describe the target in terms of warmth and competence, but also from people's own interactions with the target. These larger shifts in trait impressions may lead to more intense changes in mental state inferences. On the other hand, targets in real life may be psychologically more proximal (e.g., sharing the same physical space with the perceiver) than those used in our studies (e.g., participants only watched a video of the targets in Study 1). Prior research has shown that perceivers represent targets more concretely when they are psychologically more proximal (Ledgerwood, Trope, & Chaiken, 2010; Thornton, Weaverdyck, Mildner, & Tamir, 2019). It is likely that mental state inferences in real life are more concrete than those in our studies (e.g., situational factors exerting a greater influence). This greater gap between the higher level of abstraction in trait inferences and the even lower level of abstraction in mental state inferences may result in a smaller connection between trait and mental state inferences in real life.

The results of Studies 3 and 4 indicated that the causal effects of traits on mental state inferences were dimension-specific (Fig. 4D, bottom). Mental state inferences along the valence dimension were more sensitive to changes in perceived trait warmth than competence. Mental state inferences along the rationality dimension were more sensitive to changes in perceived trait competence than warmth. These findings are in line with prior research showing that trait warmth describes the valence of the global impressions of others across different situations (Fiske et al., 2007; Wojciszke et al., 1998); and that trait competence describes one's capability of self-control and rationality across different situations (Fiske et al., 2007; Waytz et al., 2010). On the other hand, the casual link from mental state to trait inferences was only partially dimension-specific (Fig. 3C, right). Inferences of traits along the warmth dimension were more sensitive to changes in perceived state valence than rationality and social-impact. However, the evidence for dimension-specificity regarding the effects of state rationality and state social-impact was only descriptive (Fig. 3C, right, examined columnwise). These findings reveal the nuances about how different subsets of traits and mental states are causally linked in people's mind (Fiske et al., 2007; Thornton & Mitchell, 2018; Waytz et al., 2010; Wojciszke et al., 1998).

The reasons why people rely on one unobservable variable to infer another unobservable variable are likely complex. One plausible explanation is that people have learned the reliable association between others' traits and mental states through their co-occurrence statistics in everyday life (Letzring & Adamcik, 2015). This explanation is supported by recent research showing that people learn the similarity between different mental states by observing how frequently they occur sequentially (e.g., people who feel regret usually feel angry next) (Thornton, Rmus, & Tamir, 2020). The human visual cortex also represents objects that co-occur more often in a more similar way (Bonner & Epstein, 2021). Thus, people may have learned the co-occurrences of certain traits and mental states during socialization, and then started using them to infer one another. Future research comparing participants from different younger age groups, using a longitudinal design, or analyzing individual differences would provide helpful insights (Bargh, Schwader, Hailey, Dyer, & Boothby, 2012).

Our findings have important theoretical and practical implications. Accurately identifying others' thoughts and feelings is challenging for many people in many situations (Moran et al., 2011; Poznyak et al., 2019; Shamay-Tsoory, Harari, Aharon-Peretz, & Levkovitz, 2010; Wolkenstein, Schönenberg, Schirm, & Hautzinger, 2011). Our findings on the causal link from trait to mental state inferences suggest an additional challenge: the biases and stereotypes in trait inferences might be carried over to influence mental state inferences (Mitchell, Ames, Jenkins, & Banaji, 2009). For instance, people may be more likely to attribute malicious intent to a nervous young Black man due to stereotypes of aggression (Hester & Gray, 2018), but attribute compassion to a nervous older White woman due to the stereotypes of warm and weak (Fiske, Cuddy, Glick, & Xu, 2002). Using comprehensively sampled sets of traits, mental states, and situations, our findings suggest that these carry-over biases may present in a wide range of contexts in the real world.

Conversely, our findings on the causal link from mental state to trait inferences suggest that an instance of a mental state may bias impressions of others' enduring traits. For instance, observing an emotional explosion of anger from a partner may lead to the impression that the partner may in fact be an emotional person. However, not every single instance of mental state would be informative of a person's disposition, since any person may experience uncharacteristic mental states given extreme situations. Our findings thus raise an important question: when are mental state inferences more likely to bias trait inferences? For example, when the inferred mental state is considered uncharacteristic of the target, or when the inferred mental state of the target is different from that of the perceiver, does it lead to more intense update of trait impressions or does it instead get discounted? Do certain types of mental states such as those that are negative, irrational, and socially impactful (e.g., rage, embarrassment, hate, terror) have a greater influence on trait inferences than other mental states? Existing theories such as the covariation theory (Kelley, 1973) may offer initial insights into these questions (e.g., uncharacteristic mental states likely lead to situational attribution instead of dispositional attribution). Future empirical work addressing these questions may be particularly important for understanding how social inferences impact the dynamics of close relationships.

It remains puzzling why people are predisposed to make thin-slicing trait judgments despite their inaccuracy. The facial overgeneralization hypothesis (Zebrowitz & Montepare, 2008) suggests that some physical cues were so useful for our ancestors that we overgeneralize their associations with the corresponding traits. For instance, we overgeneralize the association between features of babies (e.g., big eyes, round face) and babies' traits (e.g., warm, submissive) to baby-faced adults (i.e., baby-faced adults are perceived warm and submissive). From a functional perspective, these thin-slicing trait judgments (e.g., warmth, competence) may be crucial to guide approach or avoidance behavior for our ancestors, even though these judgments may not be always accurate in particular in the much more complex social world we face today. Our findings on the causal link from mental state to trait inferences suggest another potential explanation: people may use thin-sliced cues (e.g., facial expressions, a single behavior) to make mental state inferences, which in turn influence trait inferences. Mental state inferences based on thin-sliced cues may be accurate with respect to that particular moment. From a functional perspective, identifying others' mental states (e.g., disappointment, confusion) rapidly based on thin-sliced cues may be crucial to relationship maintenance (e.g., conflict resolution, effective cooperation). However, the utility of thin-sliced cues may not translate to trait judgments. Thus, our findings suggest another mechanism that could lead to frequently spurious trait judgments.

Several limitations constrain the conclusions drawn from this investigation. First, although we attempted to enhance the generalizability of our findings by using naturalistic stimuli (Studies 1 and 2) and sampling realistic sets of situations (Studies 3 and 4), the experimental paradigms in Studies 3 and 4 were still artificial. Generalizing to fully natural social interactions poses a significant challenge to be addressed in future research (Jolly & Chang, 2019; Nastase, Goldstein, & Hasson, 2020). Second, we also attempted to enhance the generalizability of our findings by comprehensively selecting trait and mental state terms from popular dimensional frameworks. However, this approach may lead to the omission of traits and mental states that are less optimally captured by these dimensions (e.g., the trait of "unpredictable"). Future research examining these specific traits and mental states beyond core dimensions may provide a more complete picture of the nuanced relationships between trait and mental state inferences. Relatedly, we selected our trait terms based on different dimensional frameworks across studies (see Supplementary Table S1). Tailoring the trait selection to each study's specific design allowed for more efficient capturing of the variance in trait inferences that were most relevant to the specific domain (e.g., using trait dimensions derived from face stimuli for Study 1 since participants viewed videos that featured targets' faces and other dynamic cues). However, the inconsistency in trait dimensions across studies undermined our ability to directly compare results between studies. Third, participants in all studies were paid adults, recruited online from the United States. They were not representative of the U.S. population or humans in general, a limitation shared with much psychological research (H. C. Barrett, 2020; Rad, Martingano, & Ginges, 2018). Examining the generalizability of the connection between trait and mental state inferences in more diverse populations and different cultures presents an important future direction.

Finally, we only focused on the connection between trait and mental state inferences. However, there are likely other causal factors that shape these two types of social inferences. For example, simply learning about an individual's occupation (e.g., teacher) without observing the individual's mental states might lead people to update their trait inferences about the individual (e.g., intelligent) (White & White, 2006). Learning about the situation an individual is in (e.g., getting stuck in traffic) without knowing anything about the individual's traits might still help people infer the individual's mental states (e.g., impatient). Our study only examined the relationship between trait and mental state inferences in isolation of other factors, which may not provide a complete picture of how people understand each other.

In conclusion, the present study provides new insights into how people make trait and mental state inferences in more realistic situations when artificial separations of psychological processes are removed. Perceivers use knowledge of others' traits to help make inferences of others' momentary mental states, and use knowledge of others' mental states to help make inferences of their traits. This result corroborates predictions we made in a recent theoretical paper regarding how people connect trait, mental state, and action knowledge to predict the social future (Tamir & Thornton, 2018). Our findings enrich this existing framework by showing that the connections between trait and mental state knowledge are not merely predictive, but also causal and bidirectional. Such advances in our understanding of social knowledge contribute to a cumulative, quantitative science of how people navigate the social world.

Open practices

All data, materials, and analysis code can be accessed at Open Science Framework with the following link: https://osf.io/swje9/?view_only=266eb41db0fe4ae9b9c5c41076517444.

All preregistrations can be accessed at Open Science Framework with the following links: https://osf.io/5xpng?view_only=6a0af0e751554948897cf150c139a34b; https://osf.io/6yudg/?view_only=4bcca87690fd4cd2b81fb0b05327cbb9; https://osf.io/p9m7a/?view_only=3773f7bf07bc47a39a700f735d958fbf; https://osf.io/w9jbs/?view_only=43c722dc0e2243debc5ba8ee053c2397.

Declaration of Competing Interest

None.

Data availability

All data, materials, and analysis code can be accessed at Open Science Framework with the following link: https://osf.io/swje9/?view_only=266eb41db0fe4ae9b9c5c41076517444.

Acknowledgments

We thank Diana Tamir for the situation-state co-occurrence ratings used in Studies 3 and 4. We thank Dorit Kliemann and Lindsey Tepfer for helpful input.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2023.104495.

References

Adolphs, R., Mlodinow, L., & Barrett, L. F. (2019). What is an emotion? *Current Biology: CB, 29*(20), R1060–R1064. https://doi.org/10.1016/j.cub.2019.09.008

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science, 338*(6111), 1225–1229. https://doi.org/10.1126/science.1224313

Bargh, J. A., Schwader, K. L., Hailey, S. E., Dyer, R. L., & Boothby, E. J. (2012). Automaticity in social-cognitive processes. *Trends in Cognitive Sciences, 16*(12), 593–605. https://doi.org/10.1016/j.tics.2012.10.002

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

Barrett, H. C. (2020). Towards a cognitive science of the human: Cross-cultural approaches and their urgency. *Trends in Cognitive Sciences, 24*(8), 620–638. https://doi.org/10.1016/j.tics.2020.05.007

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

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Review of Psychology, 58*(1), 373–403. https://doi.org/10.1146/annurev.psych.58.110405.085709

Bonnefon, J.-F., Hopfensitz, A., & De Neys, W. (2015). Face-ism and kernels of truth in facial inferences. *Trends in Cognitive Sciences, 19*(8), 421–422. https://doi.org/10.1016/j.tics.2015.05.002

Bonner, M. F., & Epstein, R. A. (2021). Object representations in the human brain reflect the co-occurrence statistics of vision and language. *Nature Communications, 12*(1), 4081. https://doi.org/10.1038/s41467-021-24368-2

Cannon, E. N., & Woodward, A. L. (2012). Infants generate goal-based action predictions. *Developmental Science, 15*(2), 292–298. https://doi.org/10.1111/j.1467-7687.2011.01127.x

Chen, Z., & Whitney, D. (2019). Tracking the affective state of unseen persons. *Proceedings of the National Academy of Sciences, 116*(15), 7559–7564. https://doi.org/10.1073/pnas.1812250116

Clapham, C., & Nicholson, J. (2014). *The concise Oxford dictionary of mathematics*. Oxford University Press.

DeYoung, C. G., Hirsh, J. B., Shane, M. S., Papademetris, X., Rajeevan, N., & Gray, J. R. (2010). Testing predictions from personality neuroscience: Brain structure and the big five. *Psychological Science, 21*(6), 820–828. https://doi.org/10.1177/0956797610370159

Ekman, P., & Friesen, W. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124–129.

Elsner, B., & Adam, M. (2021). Infants' goal prediction for simple action events: The role of experience and agency cues. *Topics in Cognitive Science, 13*(1), 45–62. https://doi.org/10.1111/tops.12494

Feiler, D. C., & Kleinbaum, A. M. (2015). Popularity, similarity, and the network extraversion Bias. *Psychological Science, 26*(5), 593–603. https://doi.org/10.1177/0956797615569580

Fiske, S. T., & Cox, M. G. (1979). Person concepts: The effect of target familiarity and descriptive purpose on the process of describing others1. *Journal of Personality, 47*(1), 136–161. https://doi.org/10.1111/j.1467-6494.1979.tb00619.x

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Trait as density distributions of states. *Journal of Personality and Social Psychology, 80*(6).

Foo, Y. Z., Sutherland, C. A. M., Burton, N. S., Nakagawa, S., & Rhodes, G. (2021). Accuracy in facial trustworthiness impressions: Kernel of truth or modern physiognomy? A meta-analysis. *Personality and Social Psychology Bulletin, 01461672211048110*. https://doi.org/10.1177/01461672211048110

Gweon, H. (2021). Inferential social learning: Cognitive foundations of human social learning and teaching—ScienceDirect. *Trends in Cognitive Sciences, 25*(10). Retrieved from https://www.sciencedirect.com/science/article/pii/S1364661321001789?via%3Dihub.

Hackel, L. M., Mende-Siedlecki, P., & Amodio, D. M. (2020). Reinforcement learning in social interaction: The distinguishing role of trait inference. *Journal of Experimental Social Psychology, 88*, Article 103948. https://doi.org/10.1016/j.jesp.2019.103948

Hamermesh, D. S. (2011). *Beauty pays: Why attractive people are more successful*. Princeton University Press.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450*(7169), 557–559. https://doi.org/10.1038/nature06288

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience, 7*(7), 523–534. https://doi.org/10.1038/nrn1931

Hester, N., & Gray, K. (2018). For black men, being tall increases threat stereotyping and police stops. *Proceedings of the National Academy of Sciences, 115*(11), 2711–2715. https://doi.org/10.1073/pnas.1714454115

Idson, L. C., & Mischel, W. (2001). The personality of familiar and significant people: The lay perceiver as a social–cognitive theorist. *Journal of Personality and Social Psychology, 80*(4), 585–596. https://doi.org/10.1037/0022-3514.80.4.585

Jaeger, B., Todorov, A. T., Evans, A. M., & van Beest, I. (2020). Can we reduce facial biases? Persistent effects of facial trustworthiness on sentencing decisions. *Journal of Experimental Social Psychology, 90*, Article 104004. https://doi.org/10.1016/j.jesp.2020.104004

Jamali, M., Grannan, B. L., Fedorenko, E., Saxe, R., Báez-Mendoza, R., & Williams, Z. M. (2021). Single-neuronal predictions of others' beliefs in humans. *Nature, 591*(7851), 610–614. https://doi.org/10.1038/s41586-021-03184-0

Jolly, E., & Chang, L. J. (2019). The flatland fallacy: Moving beyond low–dimensional thinking. *Topics in Cognitive Science, 11*(2), 433–454. https://doi.org/10.1111/tops.12404

Kelley, H. (1973). The processes of causal attribution. *American Psychologist, 28*(2), 107–128.

Korman, J., & Malle, B. F. (2016). Grasping for traits or reasons? How people grapple with puzzling social behaviors. *Personality and Social Psychology Bulletin, 42*(11), 1451–1465. https://doi.org/10.1177/0146167216663704

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 18*(5–6), 345–349.

Krosch, A. R., Berntsen, L., Amodio, D. M., Jost, J. T., & Van Bavel, J. J. (2013). On the ideology of hypodescent: Political conservatism predicts categorization of racially ambiguous faces as black. *Journal of Experimental Social Psychology, 49*(6), 1196–1203. https://doi.org/10.1016/j.jesp.2013.05.009

Ledgerwood, A., Trope, Y., & Chaiken, S. (2010). Flexibility now, consistency later: Psychological distance and construal shape evaluative responding. *Journal of Personality and Social Psychology, 99*(1), 32–51. https://doi.org/10.1037/a0019843

Lee, D. H., & Anderson, A. K. (2017). Reading what the mind thinks from how the eye sees. *Psychological Science, 28*(4), 494–503. https://doi.org/10.1177/0956797616687364

Letzring, T. D., & Adamcik, L. A. (2015). Personality traits and affective states: Relationships with and without affect induction. *Personality and Individual Differences, 75*, 114–120. https://doi.org/10.1016/j.paid.2014.11.011

Lin, C., Adolphs, R., & Alvarez, R. M. (2017). Cultural effects on the association between election outcomes and face-based trait inferences. *PLoS One, 12*(7), Article e0180837. https://doi.org/10.1371/journal.pone.0180837

Lin, C., Adolphs, R., & Alvarez, R. M. (2018). Inferring whether officials are corruptible from looking at their faces. *Psychological Science, 29*(11), 1807–1823. https://doi.org/10.1177/0956797618788882

Lin, C., Keles, U., & Adolphs, R. (2021). Four dimensions characterize attributions from faces using a representative set of English trait words. *Nature Communications, 12*(1), 5168. https://doi.org/10.1038/s41467-021-25500-y

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Lopes, P. N., Salovey, P., Côté, S., & Beers, M. (2005). Emotion regulation abilities and the quality of social interaction. *Emotion, 5*(1), 113–118. https://doi.org/10.1037/1528-3542.5.1.113

Lovakov, A., & Agadullina, E. (2017). Empirically derived guidelines for interpreting effect size in social psychology. *PsyArXiv.*. https://doi.org/10.31234/osf.io/2epc4

Martinez, A. M. (2019). Context may reveal how you feel. *Proceedings of the National Academy of Sciences, 116*(15), 7169–7171. https://doi.org/10.1073/pnas.1902661116

Mende-Siedlecki, P. (2018). Changing our minds: The neural bases of dynamic impression updating. *Current Opinion in Psychology, 24*, 72–76. https://doi.org/10.1016/j.copsyc.2018.08.007

Mitchell, J. P., Ames, D. L., Jenkins, A. C., & Banaji, M. R. (2009). Neural correlates of stereotype application. *Journal of Cognitive Neuroscience, 21*(3), 594–604. https://doi.org/10.1162/jocn.2009.21033

Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O'Young, D., Mavros, P. L., & Gabrieli, J. D. (2011). Impaired theory of mind for moral judgment in high-functioning autism. *Proceedings of the National Academy of Sciences, 108*(7), 2688–2692. https://doi.org/10.1073/pnas.1011734108

Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: Rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage, 222*, Article 117254. https://doi.org/10.1016/j.neuroimage.2020.117254

Oh, D., & Todorov, A. (2020, April 27). *Do regional gender and racial biases predict gender and racial biases in social face judgments? PsyArXiv*. https://doi.org/10.31234/osf.io/v7hpe

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion, 9*(1), 128–133. https://doi.org/10.1037/a0014520

Penton-Voak, I. S., Pound, N., Little, A. C., & Perrett, D. I. (2006). Personality judgments from natural and composite facial images: More evidence for a "kernel of truth" in social perception. *Social Cognition, 24*(5), 607–640. https://doi.org/10.1521/soco.2006.24.5.607

Pollet, T. V., Roberts, S. G. B., & Dunbar, R. I. M. (2011). Extraverts have larger social network layers: But do not feel emotionally closer to individuals at any layer. *Journal of Individual Differences, 32*(3), 161–169 (2011-13632-006) https://doi.org/10.1027/1614-0001/a000048 (2011-13632-006).

Poznyak, E., Morosan, L., Perroud, N., Speranza, M., Badoud, D., & Debbané, M. (2019). Roles of age, gender and psychological difficulties in adolescent mentalizing. *Journal of Adolescence, 74*, 120–129. https://doi.org/10.1016/j.adolescence.2019.06.007

Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of homo sapiens: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences, 115*(45), 11401–11405. https://doi.org/10.1073/pnas.1721165115

Rim, S., Uleman, J. S., & Trope, Y. (2009). Spontaneous trait inference and construal level theory: Psychological distance increases nonconscious trait thinking. *Journal of Experimental Social Psychology, 45*(5), 1088–1097. https://doi.org/10.1016/j.jesp.2009.06.015

Ruba, A. L., & Pollak, S. D. (2020). The development of emotion reasoning in infancy and early childhood. *Annual Review of Developmental Psychology, 2*(1), 503–531. https://doi.org/10.1146/annurev-devpsych-060320-102556

Rule, N. O., & Ambady, N. (2011). Face and fortune: Inferences of personality from managing partners' faces predict their law firms' financial success. *The Leadership Quarterly, 22*(4), 690–696. https://doi.org/10.1016/j.leaqua.2011.05.009

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion, 9*(2), 260.

Saucier, G., & Goldberg, L. R. (1996). Evidence for the big five in analyses of familiar English personality adjectives. *European Journal of Personality, 10*(1), 61–77. https://doi.org/10.1002/(SICI)1099-0984(199603)10:1<61::AID-PER246>3.0.CO;2-D

Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.00813

Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy, 2*(4), 419–436. https://doi.org/10.1207/S15327078IN0204_02

Schutte, N. S., Malouff, J. M., Bobik, C., Coston, T. D., Greeson, C., Jedlicka, C., … Wendorf, G. (2001). Emotional intelligence and interpersonal relations. *The Journal of Social Psychology, 141*(4), 523–536. https://doi.org/10.1080/00224540109600569

Shamay-Tsoory, S. G., Harari, H., Aharon-Peretz, J., & Levkovitz, Y. (2010). The role of the orbitofrontal cortex in affective theory of mind deficits in criminal offenders with psychopathic tendencies. *Cortex, 46*(5), 668–677. https://doi.org/10.1016/j.cortex.2009.04.008

Sherman, R., Rauthmann, J., Brown, N., Serfass, D., & Cooper, A. B. (2015). The independent effects of personality and situations on real-time expressions of behavior and emotion. *Journal of Personality and Social Psychology, 109*. https://doi.org/10.1037/pspp0000036

Skerry, A. E., & Spelke, E. S. (2014). Preverbal infants identify emotional reactions that are incongruent with goal outcomes. *Cognition, 130*(2), 204–216. https://doi.org/10.1016/j.cognition.2013.11.002

Smith, E. R., Miller, D. A., Maitner, A. T., Crump, S. A., Garcia-Marques, T., & Mackie, D. M. (2006). Familiarity can increase stereotyping. *Journal of Experimental Social Psychology, 42*(4), 471–478. https://doi.org/10.1016/j.jesp.2005.07.002

Sonkusare, S., Breakspear, M., & Guo, C. (2019). Naturalistic stimuli in neuroscience: Critically acclaimed. *Trends in Cognitive Sciences, 23*(8), 699–714. https://doi.org/10.1016/j.tics.2019.05.004

Spivak, M. (2008). Functions. In *Calculus* (4th ed., p. 47). Houston, Texas: Publish or Perish, Inc.

Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., & Bühner, M. (2020). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences, 117*(30), 17680–17687. https://doi.org/10.1073/pnas.1920484117

Stolier, R. M., Hehman, E., & Freeman, J. B. (2020). Trait knowledge forms a common structure across social cognition. *Nature Human Behaviour, 4*(4), 361–371. https://doi.org/10.1038/s41562-019-0800-6

Tamir, D. I., & Thornton, M. A. (2018). Modeling the predictive social mind. *Trends in Cognitive Sciences, 22*(3), 201–212. https://doi.org/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. https://doi.org/10.1073/pnas.1511905112

Teufel, C., Fletcher, P. C., & Davis, G. (2010). Seeing other minds: Attributed mental states influence perception. *Trends in Cognitive Sciences, 14*(8), 376–382. https://doi.org/10.1016/j.tics.2010.05.005

Thornton, M., Rmus, M., & Tamir, D. (2020). Mental state dynamics explain the origin of mental state concepts. *PsyArXiv.*. https://doi.org/10.31234/osf.io/kbcsj

Thornton, M., & Tamir, D. (2020). People represent mental states in terms of rationality, social impact, and valence: Validating the 3d mind model. *Cortex, 125*, 44–59. https://doi.org/10.1016/j.cortex.2019.12.012

Thornton, M. A., & Mitchell, J. P. (2018). Theories of person perception predict patterns of neural activity during mentalizing. *Cerebral Cortex, 28*(10), 3505–3520. https://doi.org/10.1093/cercor/bhx216

Thornton, M. A., Weaverdyck, M. E., Mildner, J. N., & Tamir, D. I. (2019). People represent their own mental states more distinctly than those of others. *Nature Communications, 10*(1), 1–9. https://doi.org/10.1038/s41467-019-10083-6

Thornton, M. A., Weaverdyck, M. E., & Tamir, D. I. (2019). The brain represents people as the mental states they habitually experience. *Nature Communications, 10*(1), 1–10. https://doi.org/10.1038/s41467-019-10309-7

Thornton, Mark, & Tamir, D. (2017). Mental models accurately predict emotion transitions. *Proceedings of the National Academy of Sciences, 114*(23), 5982–5987.

Todorov, A. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*(5728), 1623–1626. https://doi.org/10.1126/science.1110589

Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology, 39*(6), 549–562. https://doi.org/10.1016/S0022-1031(03)00059-3

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*(5), 675–691. https://doi.org/10.1017/S0140525X05000129

Tormala, Z. L., & Petty, R. E. (2001). On-line versus memory-based processing: The role of "need to evaluate" in person perception. *Personality and Social Psychology Bulletin, 27*(12), 1599–1612. https://doi.org/10.1177/01461672012712004

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review, 117*(2), 440–463. https://doi.org/10.1037/a0018963

Uleman, J. S., Adil Saribay, S., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology, 59*(1), 329–360. https://doi.org/10.1146/annurev.psych.59.103006.093707

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. In M. P. Zanna (Ed.)*, Vol. 28*. *Advances in experimental social psychology* (pp. 211–279). Academic Press. https://doi.org/10.1016/S0065-2601(08)60239-7.

Van Rooy, D. L., & Viswesvaran, C. (2004). Emotional intelligence: A meta-analytic investigation of predictive validity and nomological net. *Journal of Vocational Behavior, 65*(1), 71–95. https://doi.org/10.1016/S0001-8791(03)00076-9

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences, 14*(8), 383–388. https://doi.org/10.1016/j.tics.2010.05.006

White, M. J., & White, G. B. (2006). Implicit and explicit occupational gender stereotypes. *Sex Roles, 55*(3), 259–266. https://doi.org/10.1007/s11199-006-9078-z

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science, 26*(8), 1325–1331. https://doi.org/10.1177/0956797615590992

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the Spontaneousness of trait inferences. *Journal of Personality and Social Psychology, 47* (2), 237–252.

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin, 24*(12), 1251–1263. https://doi.org/10.1177/01461672982412001

Wolkenstein, L., Schönenberg, M., Schirm, E., & Hautzinger, M. (2011). I can see what you feel, but I can't deal with it: Impaired theory of mind in depression. *Journal of Affective Disorders, 132*(1), 104–111. https://doi.org/10.1016/j.jad.2011.02.010

Wood, D., Harms, P., & Vazire, S. (2010). Perceiver effects as projective tests: What your perceptions of others say about you. *Journal of Personality and Social Psychology, 99* (1), 174–190. https://doi.org/10.1037/a0019390

Wu, Y., Schulz, L. E., Frank, M. C., & Gweon, H. (2021). Emotion as information in early social learning. *Current Directions in Psychological Science, 30*(6), 468–475. https://doi.org/10.1177/09637214211040779

Xie, S. Y., Flake, J. K., Stolier, R. M., Freeman, J. B., & Hehman, E. (2021). Facial impressions are predicted by the structure of group stereotypes. *Psychological Science, 32*(12), 1979–1993. https://doi.org/10.1177/09567976211024259

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences, 45*, Article e1. https://doi.org/10.1017/S0140525X20001685

Young, L., & Saxe, R. (2009). An FMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience, 21*(7), 1396–1405. https://doi.org/10.1162/jocn.2009.21137

Zaki, J. (2013). Cue integration: A common framework for social cognition and physical perception. *Perspectives on Psychological Science, 8*(3), 296–312. https://doi.org/10.1177/1745691613475454

Zebrowitz, L. A. (2017). First impressions from faces. *Current Directions in Psychological Science, 26*(3), 237–242. https://doi.org/10.1177/0963721416683996

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass, 2*(3), 1497. https://doi.org/10.1111/j.1751-9004.2008.00109.x