

Is There a Hierarchy of Social Inferences? The Likelihood and Speed of Inferring Intentionality, Mind, and Personality

Bertram F. Malle
Brown University

Jess Holbrook
Microsoft Corporation, Redmond, Washington

People interpret behavior by making inferences about agents' intentionality, mind, and personality. Past research studied such inferences 1 at a time; in real life, people make these inferences simultaneously. The present studies therefore examined whether 4 major inferences (intentionality, desire, belief, and personality), elicited simultaneously in response to an observed behavior, might be ordered in a hierarchy of likelihood and speed. To achieve generalizability, the studies included a wide range of stimulus behaviors, presented them verbally and as dynamic videos, and assessed inferences both in a retrieval paradigm (measuring the likelihood and speed of accessing inferences immediately after they were made) and in an online processing paradigm (measuring the speed of forming inferences during behavior observation). Five studies provide evidence for a hierarchy of social inferences—from intentionality and desire to belief to personality—that is stable across verbal and visual presentations and that parallels the order found in developmental and primate research.

Keywords: attribution, theory of mind, dispositional inference, person perception, social cognition

Supplemental materials: <http://dx.doi.org/10.1037/a0026790.supp>

People have a remarkable capacity to interpret and explain human behavior. They spot a turning head and see the person's thoughts and feelings; they note a body move toward an object and know at once the person's goal; they watch two bodies step in tandem and recognize their joint intention. In making sense of myriads of human movements, people connect the observed with the unobserved—they interpret behavior by inferring mental states. This ability is essential for succeeding in the social world. Without

mental state inferences, observed behaviors look indistinct, future behaviors are difficult to predict, and communicating with others becomes utterly perplexing.

Research into the human capacity to infer mental states has been distributed over multiple disciplines such as clinical, developmental, and social psychology, cognitive neuroscience, and primatology (Baron-Cohen, Tager-Flusberg, & Cohen, 2000; Malle & Hodges, 2005; Premack & Woodruff, 1978; Saxe, Carey, & Kanwisher, 2004). Over the past 2 decades, each of these research endeavors has contributed unique findings that help elucidate how people succeed at inferring mental states (e.g., Ames, 2004), what functions these inferences serve (Bogdan, 2000; Chen, 2003; Tomasello, 1998), what brain regions might subserve them (Mitchell, 2009; Saxe et al., 2004), and how deficits in this ability affect social behavior (Baron-Cohen et al., 2000a; Frith & Corcoran, 1996). Many different inferences of mental states (understood broadly) have been studied, including a behavior's intentionality (Malle & Knobe, 1997); beliefs, desires, and intentions (Astington, 2001; Malle & Knobe, 2001; Schult, 2002); emotions and their facial expressions (Ekman, 1982; Tracy & Robins, 2008); and personality traits (Gilbert, 1989; Trope, 1986; Uleman, Saribay, & Gonzalez, 2008).

Limitations of Previous Research

Even though we have learned important characteristics of each of these inferences, past research suffered from one major limitation: Almost all studies examined a single inference at a time. Thus, little is known about how these various inferences of intentionality, mental

This article was published Online First February 6, 2012.

Bertram F. Malle, Department of Cognitive, Linguistic, and Psychological Sciences, Brown University; Jess Holbrook, Microsoft Corporation, Redmond, Washington.

The studies were partially conducted at the University of Oregon, and initial studies not reported here were part of Jess Holbrook's doctoral thesis at the University of Oregon. This work was supported by National Science Foundation Grants BCS-0937307 and BCS-0746381. We thank the many research assistants who helped with this project: Misty Auld, Hannah Boettcher, Sol Ah Han, Michelle Moore, Brian Schmidt, Lisa Sheridan, Sallie Tak, Suzy Weiss, Jasmine Yeo, and especially Holly Oh. We would also like to thank the people who provided comments, ideas, and guidance on this project: Kyle Dillon, Steve Guglielmo, Bill van Hippel, Sara Hodges, Ulrich Mayr, Steve Sloman, Jim Uleman, Ed Vogel, and Ray Vukcevic.

Correspondence concerning this article should be addressed to Bertram F. Malle, Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Box 1821, 190 Thayer Street, Providence, RI 02912.

states, and personality relate to one another. Are they made simultaneously or serially? Are they all inferred automatically, or do some have greater priority? Only studies that concurrently assess multiple inferences can answer such questions.

In addition, past research suffered from limits to generalizability that weaken conclusions even about single inferences. First, the stimulus behaviors that elicited inferences were laced with researcher-provided meaning because they were selected and heavily pretested to ensure that they elicit the desired inferences (e.g., traits, goals) in most people (Hassin, Aarts, & Ferguson, 2005; Reeder, Vonk, Ronk, Ham, & Lawrence, 2004; Winter & Uleman, 1984). A minimum of such preparation is necessary to make the stimuli relevant, but a high degree of tailoring constrains external validity. We do not know how well the constructed stimuli generalize to behaviors more broadly sampled from real life.

Second, stimulus construction was specific to the type of inference under investigation. To study trait inferences, researchers developed stimuli with trait-implicating properties; to study goal inferences, they developed stimuli with goal-implicating properties; and likewise for emotions, beliefs, and so on. Thus, we cannot compare inference types across studies because stimulus sets and inference types are confounded.

Third, behaviors were almost always presented as words or sentences. People do encounter some number of behaviors in verbal format (e.g., in newspapers, reports in conversation). However, children first come across behavior in visual ways, and human social cognition arguably evolved from exposure to visual stimuli. Some researchers have warned against generalizing from perception of text stimuli to perception of ordinary behaviors (Bassili, 1993), but few have employed visually presented dynamic behaviors in studies on social perception (Decety, Michalska, & Kinzler, 2012; Golan, Baron-Cohen, Hill, & Golan, 2006).

The Present Studies

In light of these limitations and the general goals of our investigation, we put the following demands on our studies. First, because people do not normally make only a single type of social inference, we needed to examine the unfolding of multiple inferences. Thus, our studies gave people an opportunity to make many inferences simultaneously in response to a given stimulus behavior. Naturally, we could not cover all possible social inferences. We selected four types that have been featured prominently in the literature: intentionality, desire, belief, and personality.¹

A second demand was to achieve breadth of stimulus behaviors. Thus, we included previously published tailored stimuli and added new, untailored stimuli. This way we could compare past results on single inferences from highly selected behaviors with new results on multiple inferences from a more representative range of behaviors.

A third demand was to vary modes of presentation, so we included both verbal stimuli and visual stimuli to achieve generalizability of processing characteristics and also to provide comparisons between the two dominant stimulus classes in social life.

The Simultaneous Inference Paradigm

To meet these demands, we adapted an assessment method originally developed by Smith (1984; Smith & Miller, 1983). Its basic steps were these: Teach participants single-word cues that stand for

different inference types (e.g., INTENT? for “Was the behavior intentional?”); present multiple behavioral stimuli; elicit any one of the inference types immediately after each stimulus behavior; and measure reaction times when people provide each inference.

We modified and expanded this method in the following ways: We manipulated the tailoring of stimulus behaviors; we tested verbal as well as visual stimuli; we selected a theoretically motivated set of four inference types, along with one control inference; we examined both the likelihood and speed of inferences; we assessed not only whether inferences were made but also what contents those inferences had; and (in Study 5) we examined online inferences during the dynamic unfolding of stimulus behaviors.

Hypotheses

Our primary question was whether the four major inference types (intentionality, desire, belief, and personality) may be ordered in a hierarchy with respect to their likelihood and speed. Developmental evidence suggests such a hierarchy in terms of the age of acquiring the relevant concepts: from intentionality (emerging between 6 and 18 months; Carpenter, Akhtar, & Tomasello, 1998; Woodward, 1998) to desire (emerging in the 2nd year; Wellman & Woolley, 1990), then belief (emerging in the 4th year; Wellman, Cross, & Watson, 2001) and, finally, personality (emerging not before the 6th or 7th year; Kalish & Shiverick, 2004; Snodgrass, 1976).

Evolutionary considerations favor a similar ordering. Nonhuman primates distinguish intentional from unintentional behavior (Call, Hare, Carpenter, & Tomasello, 2004), and though they can infer an agent’s desires, there is little or no evidence for genuine belief inferences (Call & Tomasello, 2008; Povinelli & Preuss, 1995). Personality inferences have not been studied.

Social psychological research is more divided. A long-dominant view places personality inferences at the forefront of social cognition (Gilbert, 1998; Ross & Nisbett, 1991; Wyer & Carlston, 1979). To illustrate, “Before forming any kind of definitive inference about others’ goals from their behaviors, it is always necessary to know something more about these individuals’ traits” (Molden, 2009, p. 38), and “a 10th of a second is often sufficient for people to make specific trait inferences” (Fiske & Taylor, 2008, p. 137). If this view is correct, then personality traits should be among the most likely and fastest to infer. On the other end of the spectrum, intentionality has been characterized as a complex judgment that requires the prior assessment of its conceptual components—primarily desire, belief, and intention (Molden, 2009; Read & Monroe, 2009).

¹ Emotion, though a frequent object of inference, was one type we did not study here. Emotion inferences uniquely challenge stimulus construction because some can be inferred easily from specific facial or bodily expressions (e.g., smiling, frowning) and specific contexts (e.g., a party, a funeral), whereas others are largely invisible (e.g., concealed annoyance or fear). Stimulus behaviors that include characteristic expressions may make inferences rather trivial, whereas behaviors that hide expressions may make inferences overly difficult. Emotion inferences may require a multifaceted approach to stimulus construction that lies beyond our present endeavor.

However, both the priority of personality trait inferences and the complex requirements of intentionality inferences can be challenged. As to trait inferences, there is evidence for a stepwise inference progression from desires and other mental states to personality (Fiedler & Schenck, 2001; Read & Miller, 2005; Reeder, 2009; Trope, 1989) and faster processing of inconsistent goal² information (Van der Cruyssen, Van Duynslaeger, Cortoos, & Van Overwalle, 2009) than inconsistent trait information (Van Duynslaeger, Van Overwalle, & Verstraeten, 2007). As to intentionality, even though people's concept of intentionality has multiple components (belief, desire, intention, etc.; Malle & Knobe, 1997), when actually determining whether a given behavior is intentional people do not normally verify each of these components (Malle, 2010). They may do so when identifying the action's specific meaning (e.g., Monroe & Reeder, 2011), but recognizing mere intentionality relies on a number of fast and simple processes: direct perception (Barrett, Todd, Miller, & Blythe, 2005; Chandler, Greenspan, & Barenboim, 1973), semantic knowledge (Jackendoff, 1990), and script activation (Read & Monroe, 2009; Schank & Abelson, 1977).

Thus, we side with the hypothesis that intentionality judgments are primary, fast, and simple and should be at least as likely and fast as desire inferences. The latter often express themselves in behavior directed to objects and are aided by cultural knowledge of canonical action–desire connections (Bruner, 1990). Next should be beliefs, which are rarely expressed in behavior (Malle, 2005), are more varied and open-ended (Ickes & Cheng, 2011), and may require context- and agent-specific knowledge. People are reluctant to make belief inferences about strangers (Malle, Knobe, & Nelson, 2007), are often slowed down when trying to infer beliefs (Apperly, Back, Samson, & France, 2008), and can be inaccurate when others' beliefs are different from their own (Barr & Keysar, 2005; Birch & Bloom, 2007; Camerer, Loewenstein, & Weber, 1989). Last come personality inferences, which are more abstract (Funder, 1991) and ambiguous (Uleman, 2005) and appear to succeed both behavior analysis (Gilbert, 1989; Trope, 1986) and goal ascriptions (Read, Jones, & Miller, 1990; Reeder, 2009).

We begin our test of this hypothesized ordering in inference likelihood and speed by using the common format of verbal stimuli (Study 1). After a replication and expansion (Study 2), we turn to a test using video stimuli (Study 3). After replicating those patterns (Study 4), we move from inference access times to online processing times (Study 5) and propose a theoretical framework that accommodates the present findings and leads to new predictions.

Study 1

Method

Participants. Thirty-nine undergraduate students participated and received course credit in return. Three participants had to be excluded; two were nonnative speakers, and one provided only 11 (out of 36 possible) valid responses. The final sample of 36 participants ranged in age from 18 to 30 years (with a median of 19) and included 70% women and 30% men. No ethnicity data were collected, but in the subject population as a whole, people self-identified as 72% White, 11% Asian-American, 5% Hispanic, 2% African American, and 10% other or undeclared.

Procedure. Up to four participants at a time arrived at the lab, met the experimenter, and completed informed consent. They first received instructions in printed form (for exact text see online supplemental material), and the experimenter read these instructions out loud while the participants read along. The instructions described the task step by step and introduced the meaning of all inference probes (e.g., that GOAL? means “Did the behavior reveal the main actor's goal?”). Participants received a reminder sheet with all the probe meanings (see Table 1), which they were free to review during the task.

Once all participants in the group indicated that they understood the task, they were randomly assigned to individual cubicles, and they sat down at a desk with a computer, keyboard, headphone/microphone set, and an instruction review sheet. All subsequent material was displayed on a CRT monitor in white font against a black background, using the *Presentation* software program (Neurobehavioral Systems, 2011). After completing all 52 trials (including practice), participants filled out a short demographic form and were debriefed.

Material. We created three sets of 12 stimulus behaviors for the current research: trait-tailored, goal-tailored, and untailored behaviors (see Appendix). We selected the trait-tailored set from Winter and Uleman (1984). These authors published attribution ratings (from “dispositional” to “totally situational”) for a number of stimulus behaviors, designed to capture how diagnostic the particular behavior was of an underlying personality trait (Winter and Uleman, 1984, p. 241). For our trait-tailored behaviors, we selected 12 of the 16 most “dispositional” behaviors. (Of the remaining four, two were unintentional and two proved too difficult to film, in anticipation of later video-based studies.)

We selected the goal-tailored set from Hassin et al. (2005). These researchers originally presented participants with several short behavior descriptions and asked them to choose one of four goals that best described the protagonist. They then selected the 20 behaviors that resulted in the highest interjudge agreement about the protagonist's goal. For our goal-tailored behaviors, we selected the 11 behaviors that were published in that article's appendix (with slight wording changes) and added one behavior of similar length and social desirability (“She sits down and turns on the tub faucet”). Because all of Hassin et al.'s (2005) goal-tailored behaviors were designed to be seen as intentional (enabling a meaningful goal inference), the trait-tailored and untailored behaviors were designed to be in principle intentional as well (see Table 2).

The untailored set was designed to be a representative group of everyday behaviors that permitted a variety of inferences. Because we planned to develop both verbal and visual stimulus behaviors, we created a large pool of both sentences and videos from multiple sources. The pool consisted of 24 videos selected from an Internet search for suitable behavior clips; 9 existing videos developed for studies on action perception (Baldwin, Baird, Saylor, & Clark, 2001) and 29 sentences that we wrote to capture everyday behaviors. The videos from the initial pool were translated into sentences

² The literature features both the terms *goal* and *desire*, but the two refer to the same type of mental state. *Goal* refers to the outcome that is represented and desired; *desire* refers to the state itself that represents the outcome. In the terminology of philosophy of mind, desires are mental acts, and goals are their contents.

Table 1
Inference Probes and Their Meanings

Probe	Question
PERSONALITY?	Did the behavior reveal a certain PERSONALITY characteristic the actor has?
GOAL?	Did the behavior reveal a certain GOAL the actor has?
THINKING?	Did the behavior reveal what the main actor was THINKING in this situation? Consider THINKING very broadly—what the actor was thinking, seeing, hearing, what s/he believed, knew, was aware of, etc.
INTENTIONAL?	Did the person INTENTIONALLY perform the behavior?
ISMALE?	Is the actor MALE?
DNTRSPND	When you see this cue, do not answer. Wait for the next screen.

for pretesting. Three judges evaluated this stimulus pool on criteria such as length, interestingness (1 = *not at all interesting* to 7 = *very interesting*), social desirability (1 = *very undesirable* to 7 = *very desirable*), intentionality (yes, no), and various scene properties. Matching the three sets on these characteristics, we created a subset of 20 untailored behaviors, which, along with the 24 tailored ones, entered another round of pretests. Samples of undergraduate students ($Ns = 20\text{--}240$) assessed the behaviors again on social desirability and interestingness and on their power to elicit certain kinds of inferences, in particular goals, beliefs, and personality traits. To capture the latter, participants were asked open-ended questions such as, “Does the behavior reveal any kind of personality characteristic the person has?” or “Does the behavior reveal a certain belief the person has?”

From the 20 untailored stimulus candidates, we selected 12 that matched the tailored ones in relevant respects (see Table 2; details in supplementary material). As designed, the trait-tailored set had higher eliciting power for trait inferences, and the goal-tailored set had higher eliciting power for goal inferences. Untailored behaviors lay in the middle between these two sets.

In addition to the 36 final stimulus behaviors, we constructed eight practice behaviors and three comparison trials asking about the actor’s gender. Further, three catch trials (to which participants were not to respond) ensured that people did not get into the habit of automatically pressing a key after each behavior, and two trials

with clearly unintentional behaviors ensured that intentionality judgments were meaningful in this task, even though all critical stimulus behaviors were reasonably interpretable as intentional.

Computer task sequence. After a brief onscreen review of instructions, participants completed the eight practice trials and, unless they had questions, proceeded to the main part of the experiment, which included the critical stimulus sentences and filler trials. This main part was divided into two blocks of 22 trials, with a short break in between.

The probe words that elicited the main inferences were: GOAL? (for desire inferences), THINKING? (for belief inferences³), PERSONALITY? and INTENTIONAL? In addition, one control probe (ISMALE?) and one catch cue (DNTRSPND) were included. The probes and their exact meaning are presented in Table 1.

As sketched in Figure 1, participants viewed each stimulus sentence for 5,000 ms, the exposure used in previous research (Winter & Uleman, 1984). Immediately thereafter, the inference probe was displayed until the participant pressed either the *Yes* key (indicating an inference) or the *No* key. Reaction times were measured from the onset of the probe to either key press. If the participant did not press a key, the probe disappeared after 3,000 ms, and the program moved on to the next trial. In the initial instructions and in the instruction reminders on screen, participants were asked to “try to answer all questions as quickly as possible.”

When participants pressed the *Yes* key in response to the GOAL? THINKING? or PERSONALITY? probe, they were immediately asked to “Please state your specific answer”—explained in the instruction as “what GOAL was revealed,” “what the person was THINKING,” and “what PERSONALITY trait was revealed.” These follow-up queries, audio recorded and transcribed, were designed to prevent participants from mere acquiescence and to encourage actual retrieval of an inference before pressing the *Yes* key. No similar follow-up queries were presented for intentionality and gender inferences or for the catch trials.

Design and analysis. The study crossed two within-subject factors: inference type (intentionality, desire, belief, personality) and behavior class (trait-tailored, goal-tailored, untailored). The 36 experimental stimulus behaviors (12 within each class) and the

Table 2
Characteristics of Stimulus Behaviors (Sentences in Studies 1 and 2) From Pretests

Characteristic	Stimulus behavior type		
	Trait tailored	Goal tailored	Untailored
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Social desirability	3.7 (1.8)	3.7 (0.9)	4.3 (0.9)
Interestingness	3.4 (0.8)	3.1 (0.7)	3.0 (0.9)
Average word count	10 (1.0)	11.5 (3.5)	10 (1.3)
Eliciting power			
For personality inferences (%)	70 (13)	44 (21)	57 (21)
For goal inferences (%)	59 (17)	83 (15)	72 (11)
For belief inferences (%)	55 (21)	35 (17)	38 (10)

Note. The mean of untailored behaviors was somewhat higher for social desirability, but not significantly so.

³ We chose the probe words GOAL and THINKING for inferences of desire and belief, respectively, because they better capture people’s linguistic use of these concepts. In other studies (Holbrook, 2006, our Study 4), we used alternative probes (e.g., BELIEF, WANTED) and found the same patterns of results.

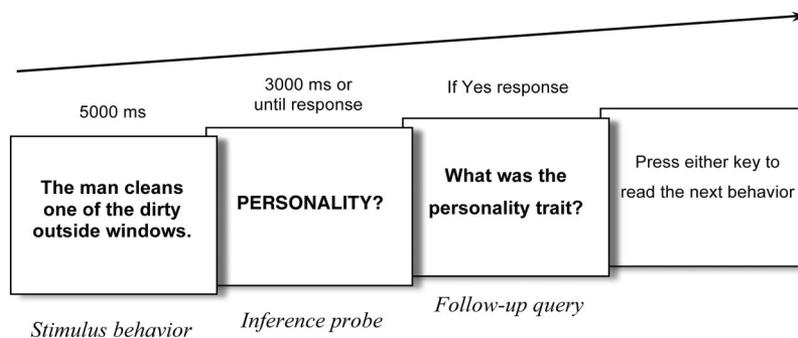


Figure 1. Event sequence for one sample trial in the computerized inference task. (In Study 1 only, the follow-up query was generic: “State your specific answer.”)

four main inference types were balanced so that each participant was probed for each inference nine times (three times from each of the three behavior classes). The ordering of stimulus sentences and inference probes was randomized within subject, with three exceptions: the 10th trial (2nd after practice) was always an unintentional behavior, the 12th trial was always a gender question, and the 14th trial was always a do-not-respond item.

Each cell of the 4×3 design had three stimulus behaviors. On this basis, likelihood of inference for each cell was the proportion of Yes (Yes + No) answers to behaviors in this cell (leaving nonresponses and excluded trials out of the denominator). Speed of inference was the average reaction time for Yes responses in that cell. We report three within-subject analyses of variance (ANOVAs). The first examines the overall patterns of likelihood and speed by inference type (four levels) for the entire set of 36 behaviors. The second conducts simple-effects tests of inference type within each behavior class (three levels) in order to examine the possibly unique patterns caused by tailoring. The inference type factor is always decomposed into three orthogonal and hierarchical contrasts: intentionality versus desire; average of intentionality/desire versus belief; and average of intentionality/desire/belief versus personality. Third, to test the specific effect of tailoring on particular inferences, we examined (a) whether desire inferences (elicited by GOAL?) varied in response to goal-tailored behaviors as compared with untailored behaviors and (b) whether personality inferences varied in response to trait-tailored behaviors as compared with untailored behaviors. For all contrasts, we report univariate t values.⁴

Invalid and corrected scores. In this and all subsequent studies, we excluded extremely short reaction times when they were located far outside the overall data distribution (typically below 400 ms) or 2.5 standard deviations below or above the particular participant’s mean. In Study 1, we excluded six trials in this way (<1% out of 1,296).

Also in all studies, we examined participants’ audio-recorded responses to follow-up queries for any obvious errors and needed corrections. When participants indicated that their Yes key press was incorrect (e.g., “no, this was not intentional”), failed to provide a verbal response after pressing the Yes key, or offered such a response more than 5 s later, the trial was reset to invalid (six trials). Conversely, when participants provided a credible and immediate verbal response but had apparently forgotten to press the Yes key or pressed the key after the verbal response, such trials were set to Yes responses but without a specific reaction time (five trials). Analyses with or

without these score corrections were highly similar, but the corrected ones are reported for precision of measurement.

Missing value replacements. Some replacements for missing reaction times were necessary when individuals who had valid reaction times for most of the 12 design cells would have been omitted from the within-subject ANOVA because of one or more cells in which they had not provided any Yes responses and, therefore, had no score of inference speed. To retain 36 participants, we used cell-based sample means to replace 57 missing values within the matrix of 432 potential values (36 participants \times 12 cells). To retain the same 36 participants in the analysis of inference likelihood, four entries were mean replaced. It should be noted that these replacements do not change the cell means of the design but somewhat lower their standard deviations.⁵ (See supplementary material for all studies’ means and standard deviations.)

Mismatched inferences. We inspected people’s verbal responses that they gave to the follow-up queries (after affirming an inference of desire, belief, or personality), and we flagged those items in which the given response clearly did not match the trial’s probe—such as when someone mentioned a desire content in a THINKING trial. There were 15% such flagged items among the critical trials. A trial-level analysis of speed data showed, however, that including or excluding those mismatched items left the results unchanged.

Results

Preliminaries. The data were reasonably well distributed, with minor positive skewness. At the person level, nine out of 36 participants showed significant (positive) skewness in their Yes reaction times, and one showed significant kurtosis. At the variable level, four out of 13 Yes reaction time aggregates (one for each cell in the design

⁴ The multilevel within-subject factors raise the issue of multivariate covariance structures (e.g., dependence among the three contrasts of the inference type factor). However, throughout the analyses reported here, no notable covariance pattern emerged, so the data are well captured by the univariate contrast tests. This lack of dependence also has theoretical relevance because it means that similarities and differences of inference types are not confounded with one another.

⁵ We also performed an alternative analysis at the trial level, treating videos as a random factor and found very similar effects as in per-person analyses. Because we aim to generalize to persons and their cognitive structures, the latter analyses are more appropriate.

as well as the gender control) showed significant skewness, and 2 showed significant kurtosis. Nonetheless, to assess the representativeness of sample means and standard errors, we computed bootstrapped means and 95% confidence intervals for the 12 cells of the design (using a macro by Andrew F. Hayes). The sample parameters were practically indistinguishable from the bootstrapped parameters.⁶

Overall pattern of inferences. (For all t values, $df = 35$.) Across all behaviors, a stable ordering of inferences emerged, both for likelihood and for speed of inference. Intentionality and desire inferences were equally likely ($M_s = 84%$ and $82%$, respectively, $t < 1$). Belief inferences ($M = 69%$) were less likely than the two ($t = 3.6$, $p < .01$). Personality inferences ($M = 46%$) were least likely of all ($t = 7.5$, $p < .001$). Intentionality and desire inferences were also similarly fast ($M_s = 1,573$ ms and $1,640$ ms, respectively, $t = 1.2$, $p > .20$). Belief inferences ($M = 1,702$ ms) were slower than the two ($t = 2.02$, $p = .05$). Personality inferences ($M = 1,797$ ms) were slowest of all ($t = 3.1$, $p < .01$). Put differently, the linear effect (from intentionality to desire to belief to personality) accounted for 41% of the total variance in inference likelihood (with 5% nonlinear effects) and for 14% of the total variance in inference speed (<1% nonlinear effects).

Likelihood of inferences within behavior classes. As Figure 2 shows, patterns of inference likelihood for untailored and goal-tailored behaviors were very similar. Within each, intentionality and desire inferences were equally likely (for untailored, $t < 1$, for goal-tailored, $t = 1.1$, both $ps > .25$), belief inferences were less likely than the two (for untailored, $t = 2.7$, for goal-tailored, $t = 3.9$, both $ps \leq .01$), and personality inferences were far less likely than the rest (for untailored, $t = 6.8$, for goal-tailored, $t = 7.7$, both $ps < .01$). This pattern washed out for trait-tailored behaviors, where intentionality inferences were somewhat more likely than the average of all other inferences ($t = 2.0$, $p = .06$), which themselves did not differ from one another ($ts < 1$).

Speed of inferences within behavior classes. Figure 3⁷ displays the speed results for the full design. Serving as a comparison standard, inferences about gender showed an average access speed of 1,585 ms (saying *Yes* to the question MALE?).

In response to untailored behaviors, intentionality inferences ($M = 1,490$ ms) were marginally faster than desire inferences ($M = 1,592$ ms, $t = 1.6$, $p = .12$), belief inferences ($M = 1,749$ ms) were slower than the two ($t = 2.5$, $p < .05$), and personality inferences ($M = 1,935$ ms) were the slowest of all ($t = 4.2$, $p < .01$).

In response to goal-tailored behaviors, there was less differentiation, with intentionality inferences, desire inferences, and belief inferences statistically indistinguishable, and only personality inferences ($M = 1,835$ ms) slower than the rest ($M = 1,653$ ms, $t = 2.1$, $p < .05$). Inspecting the pattern of means in Figure 3 shows that the lower differentiation in goal-tailored behaviors, compared with untailored behaviors, is primarily due to a slowing of intentionality inferences.

In response to trait-tailored behaviors, finally, differentiation was the weakest. Only inferring intentionality ($M = 1,565$) was faster than inferring desire ($M = 1,730$, $t = 1.9$, $p = .06$).

Specific tailoring effects. Goal-tailored behaviors had no effect on desire inferences, for either likelihood or speed ($ts < 1$). By contrast, trait-tailored behaviors had a strong effect on personality inferences. They increased their likelihood to 70%, compared with

36% in response to untailored behaviors ($t = 4.4$, $p < .01$), and they sped them up to 1,620 ms, compared with 1,935 ms in response to untailored behaviors ($t = 5.1$, $p < .01$). Interestingly, trait-tailored behaviors inhibited desire inferences to some extent, reducing their likelihood from 86% to 72% ($t = 2.3$, $p < .05$) and slowing them down from 1,592 ms to 1,730 ms ($t = 1.9$, $p = .06$).

Discussion

Inference hierarchy. Study 1 yielded evidence for a hierarchy of social inferences. Across behavior classes, people were most likely to infer intentionality and desire and inferred them the fastest; they were less likely to infer beliefs, and if they did, those inferences were slower; and they were least likely to infer personality, and if they did, those inferences were the slowest. This ordering runs parallel to that found in the evolutionary and developmental literature. In particular, we confirmed the consistently greater difficulty and greater processing demands of belief inferences over desire inferences and the greater difficulty and processing demands of personality inferences over mental state inferences. No previous investigation had considered jointly all four inference types; their ordering in simultaneous operation is of greatest interest here.

Gender. Participants could read off gender from the stimulus phrases “A women” and “A man.” Thus being a rather minimal inference, the gender retrieval speed of 1,430 ms to 1,580 ms provides a lower bound for the retrieval of any other (more complex) inference from the stimulus behaviors. In this light, the speed of accessing intentionality and desire inferences within 1,500 ms to 1,600 ms (in response to untailored behaviors) is impressive.

Behavior effects. In addition to the hierarchical pattern of inference types, we also found effects due to stimulus behaviors—more precisely, due to the selection of behaviors that favor particular inferences. Untailored behaviors showed a clean linear ordering among inferences for both likelihood and speed: intentionality, followed by desires, beliefs, and personality. Goal-tailored behaviors maintained this pattern for likelihood of inferences and largely for speed, though intentionality judgments became somewhat slower to access than desire inferences. Trait-tailored behaviors washed out this ordering for both likelihood and speed, in part by facilitating and speeding up personality inferences and in part by hindering and slowing down desire inferences.

Some impact of behavior classes was of course expected because of the purposeful design and selection of inference-tailored behaviors. However, the tailored behaviors exerted less facilitation for their “targeted” inferences than they exerted inhibition of other inferences. Goal-tailored behaviors did not facilitate desire inferences but inhibited the speed of intentionality inferences; trait-

⁶ For likelihood scores, sample means deviated on average by +0.1% (positive numbers indicate larger sample parameters), with deviations for individual cell means ranging from -0.5% to +0.7%. Sample confidence interval widths differed by +0.4% (-1.2% to +1.3%). For speed scores, sample means deviated on average by -1 ms (-5 ms to +3 ms), and the average deviation of confidence interval widths was 0 ms (from -20 ms to +16 ms).

⁷ For ease of distinction, likelihood results are displayed in horizontal bar format, and reaction time results are displayed in column format.

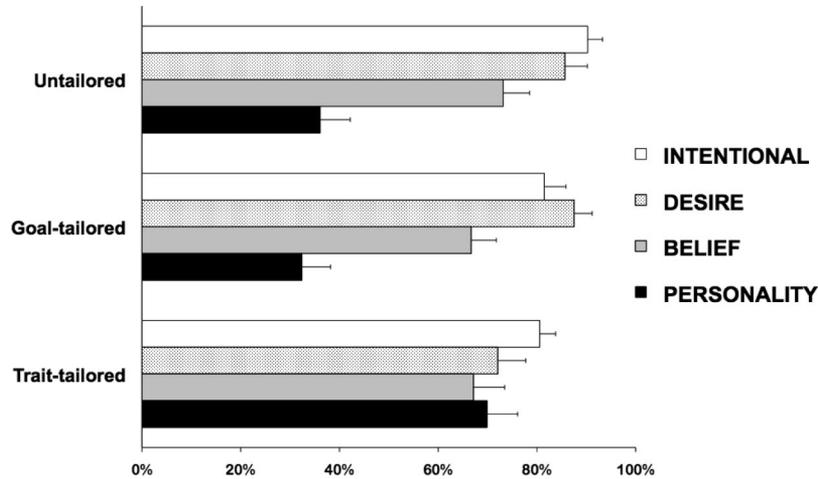


Figure 2. Likelihood of making four types of inference within three classes of behavior in Study 1. Error bars indicate within-cell standard errors.

tailored behaviors facilitated personality inferences but also inhibited the speed of desire inferences. We reserve speculation about the reason for such inhibition effects until after their consistency has been established in the following studies.

An important point worth noting is that the linear inference hierarchy seen for the 12 untailored behaviors closely mirrored the linear inference hierarchy seen for the entire set of 36 behaviors. Thus, whatever facilitating and inhibiting effects goal-tailored and trait-tailored behavior had, they largely canceled each other out, and untailored behaviors seem to represent “behaviors in general” very well.

Alternative accounts. We now consider a number of possible explanations of our results that rely on simpler assumptions than actual processing differences between inference types. A first possibility is that the differences in inference speed are driven entirely by differences in inference likelihood (and the latter may be vulnerable to other challenges; see below). However, even though the overall patterns of likelihood and speed show substantial similarities, the variables as measured are hardly related. The overall trial-by-trial correlation between responding *Yes* (vs. *No*)

and the latency of that response was $r(1,639) = .002$. Among untailored behaviors, it was $r(386) = -.07$; among goal-tailored behaviors, $r(219) = -.06$; among trait-tailored behaviors, $r(365) = -.17$. Broken down by inference type, for intentionality inferences, it was $r(379) = .02$; for desire inferences, $r(322) = -.25$; for belief inferences, $r(301) = -.18$; and for personality inferences, $r(328) = .12$.

A second possibility is that presenting mostly intentional stimulus behaviors may have biased the results in favor of intentionality and desire inferences and against other inferences. For likelihood data, this is partially correct. Intentional behaviors enable and therefore favor the likelihood of making intentionality and desire inferences. But there is no reason to believe that intentional behaviors *inhibit* the likelihood of, say, making personality inferences. After all, Winter and Uleman’s (1984) trait-tailored behaviors were almost exclusively intentional, following Jones and Davis’s (1965) claim that people draw personality inferences predominantly from intentional behaviors.

As to the speed of inferences, studying intentionality, desire, and other inferences in their simultaneity necessarily requires

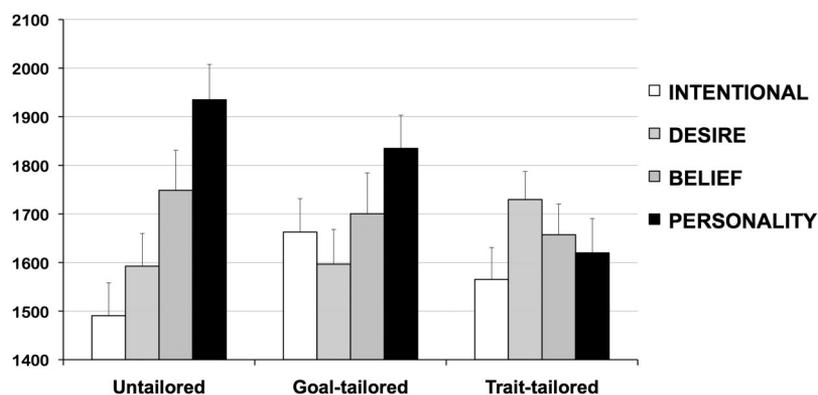


Figure 3. Speed of making four types of inference within three classes of behavior in Study 1. Error bars indicate within-cell standard errors.

presenting intentional behaviors as stimuli so that intentionality and desire inferences are actually made and can be compared with belief and personality inferences. Most important, speed data are orthogonal to the lower overall likelihood of belief and personality inferences because inference speed was measured only for those items in which people did in fact make the relevant inference.

We also examined a more specific version of the above concern: that the speed of intentionality inferences in particular may have been aided by perceivers' growing recognition of the high base rate of intentional behaviors. Encountering many intentional behaviors, however, might in fact demand more careful judgment in order to weed out the few behaviors that are not intentional. Indeed, analyzing intentionality inferences at the trial level, we found that rejecting the intentionality of unintentional behaviors was actually faster ($M = 1,359$ ms) than affirming the intentionality of intentional behaviors ($M = 1,548$ ms), $F(1, 351) = 11.6$, $p < .001$, so people did not show fast intentionality judgments because of an endorsement practice effect. Moreover, other studies in our lab have used a stimulus set containing an additional 10 unintentional behaviors along with the 24 untailored and goal-tailored behaviors. In this context of one-third unintentional behaviors, intentionality judgments showed an average speed of 1,450 ms, which was even faster than we found in Study 1.

Study 2

In Study 2, we aimed at replicating the above results while improving methodological details. We asked participants to rehearse the meaning of inference probes; we slightly adjusted the inference probes themselves; we used balanced forms rather than randomized trial orders; and we sharpened the way the queries following *Yes* responses were introduced and formulated.

Method

Participants. Fifty-eight undergraduate students participated and received course credit in return. One did not complete the experiment because of a computer failure; two did not follow instructions (giving verbal responses instead of *Yes/No* key presses); four were confused by the instructions (due to experimenter error); and four were nonnative speakers who had trouble with the instructions and the verbal stimuli. The remaining 47 participants ranged in age from 18 to 69 years ($Mdn = 19$), and 64% were women and 36% were men.

Procedure. The procedure was identical to that of Study 1.

Material and task. The following task improvements were made over Study 1. First, we inserted a prepractice phase between the verbal instructions and the eight practice trials to further familiarize participants with the meaning of the six inference probes. During this phase, a probe appeared on the screen and participants had to say out loud what it meant, after which the screen showed the correct meaning. Upon completing this prepractice for all probes, the actual stimulus practice began. Second, we slightly reformulated the inference questions to address participants more directly: "Did you detect?" rather than "Did the behavior reveal?" We also shortened the definition of the THINKING probe: "Did you detect what the main actor was THINKING (was aware of, knew, saw, etc.) in this situation?" We

turned the shortest probe word "GOAL?" into "THEGOAL?" to bring it closer in length to the other probe words. Third, we improved the follow-up queries to *Yes* responses. The intentionality inference received a follow-up query as well, the instructions for all queries were more detailed, and the onscreen queries were more explicit: "What behavior did you judge as intentional?" "What was the goal?" "What was the main actor thinking, aware of, etc.?" "What was the personality trait?" Fourth, randomizing the order of stimulus behaviors and inference probes in Study 1 caused occasional quadruplets of trials with the same inference probe (e.g., four times GOAL? in a row). To avoid such patterns we randomly assigned participants in Study 2 to one of four balanced forms. Within each form, each inference type was paired with an equal number of stimulus sentences from each behavior class (untailored, goal-tailored, trait-tailored); and across forms, each sentence was paired exactly once with each inference type. Moreover, none of the forms contained more than two successive trials featuring the same inference type or behavior class.

Exclusions, corrected scores, and replacements. Following the guidelines outlined in Study 1, we excluded three outliers and, after analysis of audio files, eliminated nine trials and corrected five (<1% of 1,692 trials). To retain participants in the within-subject analyses, we entered 49 mean-replaced values into the matrix of 564 scores.

Mismatched inferences. The probe practice, improved instructions, and more explicit follow-up queries in Study 2 reduced mismatches between probed inference type and verbal response to 11% (compared with 15% in Study 1). Two thirds of these mismatches consisted of participants mentioning a desire content, even though a desire was not probed. The most susceptible probes were INTENTIONAL and THINKING (13% of their relevant trials had a desire content). As in Study 1, results did not change as a function of such mismatched items.

Results

Overall pattern of inferences. (For all t values, $df = 46$.) The pattern of inferences across all behaviors closely replicated that of Study 1. Intentionality and desire inferences were equally likely ($M = 91%$, $t < 1$). Belief inferences were less likely than the two ($M = 69%$, $t = 6.9$, $p < .001$). Personality inferences were least likely of all ($M = 55%$, $t = 8.8$, $p < .001$). Likewise, intentionality inferences and desire inferences were similarly fast ($M_s = 1,535$ ms and 1,590 ms, respectively, $t = 1.5$, $p > .13$). Belief inferences were slower than the two ($M = 1,754$ ms, $t = 5.0$, $p < .001$). Personality inferences were slowest of all ($M = 1,820$ ms, $t = 5.4$, $p < .001$). The linear effect explained 50% of the total variance in the likelihood data (5% nonlinear effects), and it explained 30% of the total variance in the speed data (1% nonlinear effects).

Likelihood of inferences within behavior classes. As Figure 4 shows, likelihood patterns within behavior classes were very similar to those in Study 1. In response to both untailored and goal-tailored behaviors, intentionality and desire inferences were equally likely ($p_s > .25$), belief inferences were less likely ($t_s > 5.1$, $p_s < .01$), and personality inferences were less likely yet ($t_s > 7.9$, $p_s < .01$). In response to trait-tailored behaviors, intentionality and desire inferences were still equally likely ($t < 1$), belief inferences were less likely ($t = 4.0$, $p <$

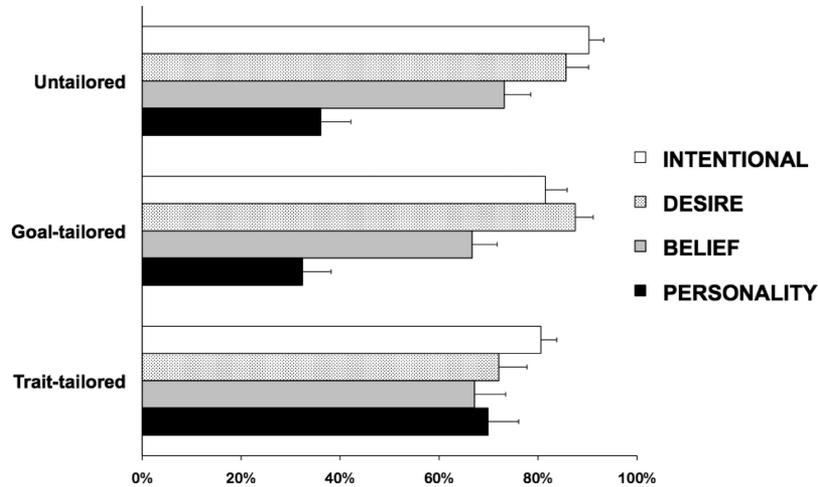


Figure 4. Likelihood of making four types of inference within three classes of behavior in Study 2. Error bars indicate within-cell standard errors.

.01), but personality inferences no longer differed from the average of the other three inferences ($p > .20$).

Speed of inferences within behavior classes. Serving as a comparison standard, inferences about gender showed an average access speed of 1,667 ms (saying *Yes* to the question MALE?).

As Figure 5 shows, in response to untailored behaviors, desire inferences ($M = 1,537$ ms) and intentionality inferences ($M = 1,477$ ms) were equally fast ($t < 1$), belief inferences ($M = 1,683$ ms) were slower than the two ($t = 3.0, p < .01$), and personality inferences were the slowest of all ($M = 1,785$ ms, $t = 4.1, p < .01$). In response to goal-tailored behaviors, intentionality inferences ($M = 1,606$ ms) slowed down relative to desire inferences ($M = 1,486$ ms, $t = 2.1, p < .05$); belief inferences ($M = 1,699$ ms) were slower than the two ($t = 2.9, p < .01$), and personality inferences ($M = 1,931$ ms) were slower yet ($t = 3.6, p < .01$).

In response to trait-tailored behaviors, intentionality inferences ($M = 1,520$ ms) were faster than desire inferences ($M = 1,746$ ms, $t = 3.6, p < .01$), belief inferences ($M = 1,878$ ms) were slower than the two ($t = 4.1, p < .01$), but personality inferences ($M =$

1,743) were no longer slower than the average of the other inferences ($t = 1.6, p = .12$). Descriptively, the best way to characterize this pattern is that intentionality inferences were faster than all other inferences, and the latter were compressed in the slower range and statistically indistinguishable ($ps > .12$).

Specific tailoring effects. Goal-tailored behaviors again showed no effect on desire inferences, for either likelihood or speed ($ts < 1$). Trait-tailored behaviors had a strong effect on personality inferences for likelihood (+38% points; $t = 6.6, p < .01$) but, unlike in Study 1, not for speed ($t < 1$). Once more, trait-tailored behaviors inhibited desire inferences, reducing their likelihood by 8.5% ($t = 1.95, p = .06$) and slowing their speed by 209 ms ($t = 3.6, p < .01$). They also slowed down belief inferences by 195 ms ($t = 2.8, p < .01$).

Discussion

The results of Study 2 mirrored those of Study 1. The inference pattern across all behaviors and within untailored behaviors duplicated the hierarchy from intentionality to personality inferences.

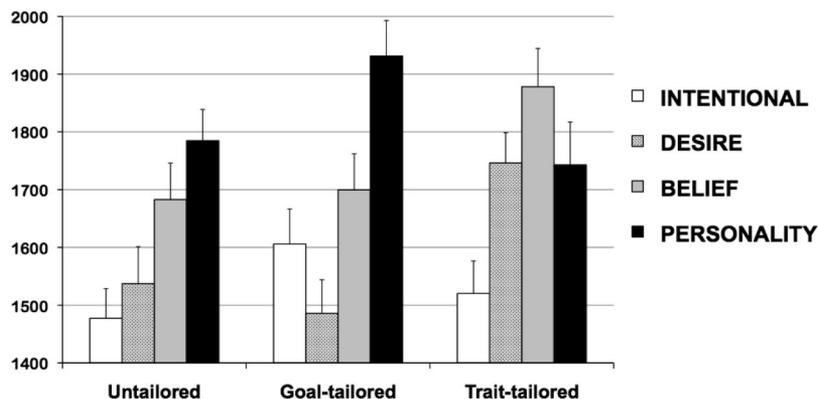


Figure 5. Speed (in ms) of making four types of inference within three classes of behavior in Study 2. Error bars indicate within-cell standard errors.

Variations due to tailoring also largely replicated. Goal-tailored behaviors elicited an ordering similar to untailored behaviors but, as in Study 1, slowed down intentionality inferences. Trait-tailored behaviors compressed desire, belief, and personality inferences to a similar speed, all slower than intentionality inferences. The main difference from Study 1 was a slightly stronger boost of likelihood, but not speed, for personality inferences in response to trait-tailored behaviors (compared with untailored behaviors).

The small changes to probes and definitions in Study 2 had no measurable impact, which suggests that the same inferences can be elicited in multiple ways—asking either about object properties (“Did the behavior reveal?”) or about subjective perception (“Did you detect?”). The change from GOAL to THEGOAL also had no marked effect. If anything, it further sped up reaction times, ruling out the possibility that the short probe word in Study 1 unduly favored people’s speed of access to desire inferences.

Tailored behaviors again showed more inhibiting than facilitating effects. Goal-tailored behaviors did not facilitate desire inferences but inhibited somewhat the speed of intentionality inferences. Trait-tailored behaviors did not facilitate the speed (only the likelihood) of personality inferences but inhibited the speed of both desire and belief inferences.

One possible reason for these inhibition effects is that the tailoring process impoverishes stimuli, making one aspect of the behavior so salient that other aspects become difficult to discern. However, people were still able to make the relevant inferences (e.g., desire inferences for trait-tailored behaviors), only at slightly lower rates. The greater salience of tailored stimulus aspects may not preclude inferences but may push them into the back of the perceiver’s mind, slowing access times. Whatever potential impoverishment there may be, our hope was that video stimuli, employed in the subsequent studies, would weaken these effects.

We also note that the inhibiting effects are largely probe-specific—goal-tailored behaviors slowed down only intentionality inferences, and trait-tailored behaviors consistently slowed down only desire inferences. One feature that characterizes all 12 goal-tailored behaviors is that the goal object is always implied rather than mentioned in the sentence. Moreover, nine of these 12 goal objects are actions (e.g., buying, going to bed, washing the car). Thus, participants parse one action that is described and another action that is implied. Even if they encode both, when asked INTENTIONAL? they must decide which “counts” as the primary one, and this decision may slow them down. As to trait-tailored behaviors, inspection of the reaction time patterns for all 12 behaviors suggested that seven were right in the range of untailored behaviors (1,400–1,700 ms) but five inhibited access to desire inferences (reaction times of 1,800–1,900 ms). These inhibiting behaviors appeared to be either of questionable intentionality, according to the *Yes* rates for intentionality (picking teeth, stepping in front of the line, solving a mystery) or were unusual (the plumber slipping an extra \$50 into his wife’s purse).

Despite such behavior-specific variations, the hierarchy of inferences held strongly when averaging over the entire set of 36 behaviors. Taking this set as spanning the breadth of behaviors in the real world, we can be reasonably confident in the generalizability of this result.

Study 3

Our next goal was to examine simultaneous inferences of intentionality, desire, belief, and personality from dynamic visual stimuli. This is not merely a replication in an often recommended but rarely used presentation medium. Inferences from visually presented behaviors most naturally model the inferences people make while interacting in the social world (Freeman & Ambady, 2011). Moreover, visual displays should trigger the evolutionarily and developmentally oldest inferential processes—those that do not require linguistic framing of the stimulus. Before mastering language, infants have to make sense of the behaviors unfolding before them, and early hominids faced the same challenge. Converging evidence across verbal and visual presentation modes would assure broad generalizability and provide credible support for a genuine hierarchy among social inferences.

To underscore this point, we highlight three differences between verbal and visual stimuli. First, sentences provide stimulus–response compatibility in that both behavior and inference are in a linguistic format; in the case of visual stimuli, the perceiver must transform visual information into linguistic responses. Second, many verbal behavior descriptions provide the content of the intentional action, and sometimes of the goal, in their verb phrases (e.g., “The butcher *writes a letter* to the editor about air pollution”; “The accountant *takes the orphan to the circus*”). When semantic processing already reveals intentionality, the likelihood and speed of inferring intentionality will naturally be fast. Such an aid is absent in the case of video stimuli because the action has to be extracted from the continuous motion stream (Baldwin & Baird, 2001). Likewise, sentences that were tailored to elicit personality inferences provide word meanings and associations that may require little actual inferential activity (e.g., “The successful filmmaker *gives his ailing mother \$20 a month*”). Here, too, videos demand more processing to arrive at the same inference. Third, the information density of verbal stimuli (in this and previous studies) has been kept relatively low by limiting length (around 10 words) and syntactic complexity (1–2 clauses). Dynamic visual stimuli are likely to be more complex in these respects, though there is no familiar metric that would quantify this information density.

Method

Participants. Of the original sample of 47 participants, seven did not yield usable data: three because of computer breakdowns; two because they offered 13 or fewer *Yes* responses (out of 36 critical trials); and two for failing to follow instructions. Forty participants remained.

Materials. Six of 12 untailored stimuli had originally been selected as video clips (and were translated into sentences for Studies 1 and 2), and these videos were used directly. For the remaining stimuli (six untailored, 12 goal-tailored, 12 trait-tailored) we created novel video versions. Each video was to depict its sentence stimulus as closely as possible, including the number of characters, their social categories (e.g., gender, age), their actions revealed in the sentence verbs, and the setting (e.g., living room, restaurant, campus). We could not represent Winter and Uleman’s (1984) profession labels (e.g., tailor, accountant), and in a few cases, the setting had to be a synecdoche (e.g., *exiting a taxi outside the airport* for “The man with the luggage goes to Den-

ver”; *welcoming a guest at the door* for “The professor has his new neighbor over for dinner.”). All actors were amateurs recruited from the authors’ social networks.

All but two videos were recorded with a digital video camera, edited with iMovie software on a Macintosh computer, and converted to .mpg video format to be presented on a personal computer. Their size was 720×480 pixels. The two remaining videos had been downloaded from the Internet, edited in the same way, and their size was 320×240 pixels. As the depicted behaviors varied in complexity, the videos’ length ranged accordingly from 4 to 12 s, with an average of 7 s (and equal mean length across the three behavior classes). Sixteen of the videos had editing cuts (three untailed, five trait-tailored, and eight goal-tailored behaviors). All but four videos had an audio track (ambient sounds or dialogue), and two had a voiceover to reveal the text of a letter. All videos, along with a tabulation of features for each, are available in the online supplemental material.

Procedure. The procedure was identical to that in Studies 1 and 2, except that participants responded to inference probes immediately after the end of each video. Videos were displayed on a 17 in. (43.18 cm) CRT monitor at about 25 in. (63.5 cm) from the participant, resulting in visual angles (image width) of about 17° (for the larger videos) and 9° (for the two smaller videos).

Exclusions, corrections, replacements. Following procedures described in Study 1, we excluded one outlier, eliminated 13 invalid trials, and corrected four trials (out of 1,440 critical trials). To retain participants in the within-subject analyses, we entered 27 mean-replaced values into the matrix of 520 scores.

Mismatched inferences. Only 5% of verbal responses to follow-up queries were inconsistent with the probed inference (e.g., a desire description for a THINKING probe), which was considerably lower than in Study 2. Excluding these trials did not alter the results.

Results

Preliminaries. Inferences from video stimuli in this study were accessed more quickly ($M = 1,376$ ms) than inferences from text

stimuli in Study 1 ($M = 1,695$) and Study 2 ($M = 1,707$). Providing a benchmark, the speed of accessing gender inferences was 1,267 ms.

Overall pattern of inferences. (For all t values, $df = 39$.) Across behaviors, a stable ordering of inferences emerged again, both for likelihood and speed of inference. Intentionality and desire inferences were comparably likely ($M_s = 85\%$ and 89% , respectively; $p > .10$) and more likely than belief inferences ($M = 72\%$, $t = 3.9$, $p < .001$), which were themselves more likely than personality inferences ($M = 42\%$, $t = 6.7$, $p < .001$). Intentionality and desire inferences were also fastest ($M_s = 1,332$ ms and $1,304$ ms, respectively, $t < 1$), belief inferences were slower ($M = 1,418$ ms, $t = 2.9$, $p < .01$), and personality inferences were slower yet ($M = 1,525$ ms, $t = 2.6$, $p = .01$). A linear effect term explained 46% of the total variance in the likelihood data (13% nonlinear effects) and 19% of the total variance in the speed data (5% nonlinear effects).

Likelihood of inferences within behavior classes. As Figure 6 shows, the pattern of inference likelihoods in video-based Study 3 was highly similar to that of text-based Studies 1 and 2. In response to untailed and goal-tailored behaviors, intentionality and desire inferences were highly likely and indistinguishable ($p_s > .15$), belief inferences were less likely ($t_s > 2.4$, $p_s < .02$), and personality inferences were less likely yet ($t_s > 6.9$, $p_s < .01$). In response to trait-tailored behaviors, intentionality and goal inferences were equally likely ($t < 1$), belief inferences were still somewhat less likely ($t = 2.1$, $p < .05$), but personality inferences were now indistinguishable from belief inferences, as in Study 1.

Speed of inferences within behavior classes. As Figure 7 shows, untailed behaviors elicited equally fast intentionality inferences ($M = 1,276$ ms) and desire inferences ($M = 1,300$ ms, $t < 1.0$). Belief inferences ($M = 1,434$ ms) were slower than the two ($t = 2.9$, $p < .01$), and trait inferences were the slowest of all ($M = 1,582$ ms, $t = 3.4$, $p < .01$). Goal-tailored behaviors slowed down intentionality inferences ($M = 1,404$ ms) compared with desire inferences ($M = 1,242$ ms, $t = 2.5$, $p < .05$), belief inferences ($M = 1,343$ ms) were indistinguishable from the two

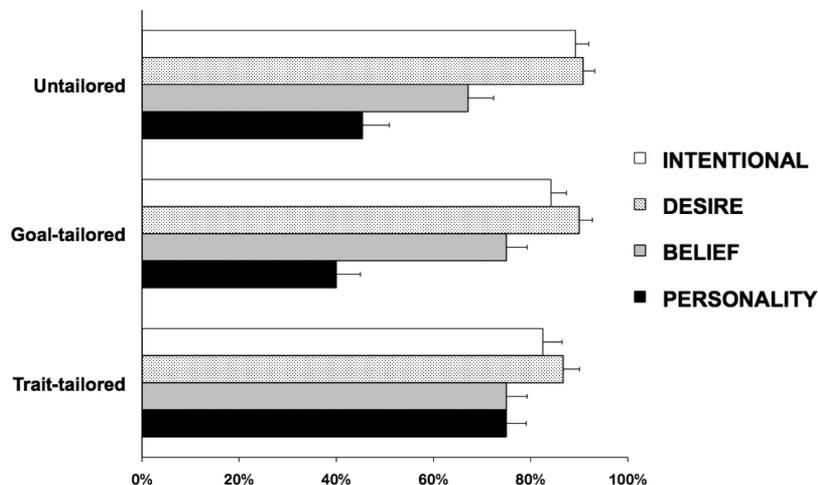


Figure 6. Likelihood of inferring intentionality, desire, belief, and personality (video stimuli, Study 3). Error bars indicate standard error.

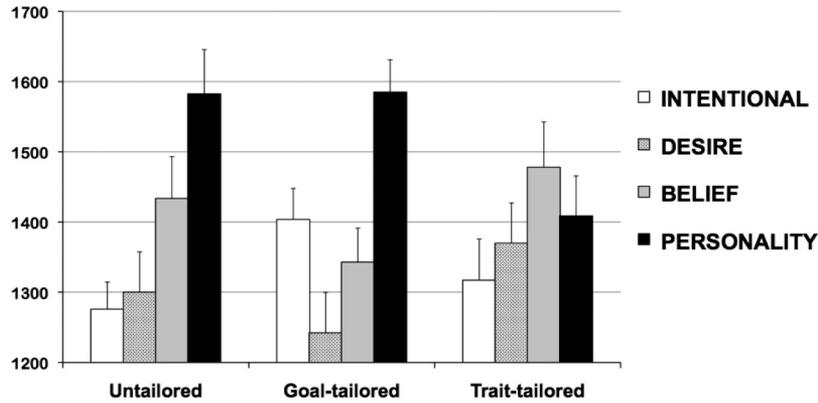


Figure 7. Speed of inferring intentionality, desire, belief, and personality (video stimuli, Study 3). Error bars indicate within-cell standard errors.

($t < 1$), and personality inferences ($M = 1,585$ ms) remained slowest of all ($t = 5.1, p < .01$).

Trait-tailored behaviors again compressed response speeds. Intentionality inferences ($M = 1,317$) did not differ from desire inferences ($M = 1,370, t < 1$), and though belief inferences ($M = 1,478$) were somewhat slower than the two ($t = 2.1, p < .05$), personality inferences ($M = 1,409$) did not differ from the other three inference types ($t < 1$).

Specific tailoring effects. There was again no reliable tailoring effect for desire inferences in either likelihood or speed ($t_s < 1.2$). The familiar tailoring effect for personality inferences was reliable for both likelihood (+30%; $t = 5.7, p < .01$) and speed (−174 ms, $t = 2.2, p < .05$). Desire inferences were slightly slowed by trait-tailored behaviors (+70 ms, $p > .20$).

Discussion

Study 3 used video stimuli to elicit the same set of social inferences as in Studies 1 and 2, and the patterns of both likelihood and speed, across and within behavior classes, replicated the previously found patterns. For inference likelihood, a well-differentiated hierarchy held in response to both untailored and goal-tailored behaviors, such that people made intentionality and desire inferences for almost all behaviors, belief inferences for about three fourths of behaviors, and personality inferences for no more than half of behaviors. Likewise, the patterns of inference speed of Studies 2 and 3 are almost indistinguishable (see Figures 5 and 7). Intentionality and desire inferences are equally fast, belief inferences are slower than desire or intentionality inferences, and personality inferences are generally the slowest of all.

Tailoring effects were still present with video stimuli, though somewhat weaker than with text-based stimuli. Trait-tailored behaviors again leveled the speed differences among inference types that naturally existed in response to the other behavior classes. But, unlike in Studies 1 and 2, they did this more by speeding up the normally slower personality inferences and less so by slowing down the normally faster other inferences.

A point worth noting is that accessing an inference about gender appears to be no faster than accessing an inference about intentionality or desires; in fact, the pattern of means suggested that some intentionality inferences (from untailored behaviors)

and some desire inferences (from goal-tailored behaviors) tended to be retrieved faster than gender inferences, though none of the comparisons were statistically significant. This comparison is particularly noteworthy for desire inferences because retrieval for both gender and intentionality need only deliver one of two values (yes or no) whereas desire inferences have a specific content.

The equality of results across the text- and video-based stimuli is remarkable, as many researchers would have predicted visual stimuli to show a distinct pattern (e.g., Amit, Algom, & Trope, 2009) or at least to create more variance in participants' responses due to the complexity of the stimulus. But before we draw too strong conclusions, we need to replicate this similarity and generalize the findings to alternate inference probes.

Study 4

Studies 1–3 employed a relatively constant set of inference probes, so one might worry that the consistency of results is bound to these probes. Initial evidence against such an interpretation comes from Holbrook (2006), who had used the words INTENT? BELIEF? and TRAIT? and found generally similar patterns as we have shown in Studies 1–3. But there were ambiguities associated with at least two of these early probes,⁸ so we wanted to test alternative probe words that did not carry known problems of interpretation. Following theoretical expectations, we developed three such probes.

The first concerned desire inferences. Responses to THE-GOAL? may have been easier because the perceiver could get by with attending to desired objects in the world (goal = a puppy, a cleaner environment) rather than the agent's mental state of desire that represents that object. We therefore asked directly about the agent's

⁸ Follow-up responses to INTENT? contained more than 50% goal contents, so we cannot be certain that people had made intentionality inferences rather than goal inferences in response to this probe. (After the switch to INTENTIONAL? we greatly reduced these mismatches.) The problem with the BELIEF? probe was that some people interpreted it as a stable disposition, indicating strongly held "beliefs" or values. Accordingly, this dispositional reading produced reaction times that were indistinguishable from those of personality inferences.

desire state with the probe WANTED? The second change concerned belief inferences. Responses to THINKING? may have shown a lower likelihood because thinking could be interpreted as a complex mental process, not merely a belief about some event in the world. We targeted more directly a representational state with the probe AWAREOF?⁹ Finally, the probe for gender inferences was changed from ISMALE? to FEMALE? Probes for intentionality and personality inferences remained the same.

Method

Participants. Undergraduate students who were enrolled in introductory psychology or linguistics courses participated in exchange for partial course credit. Of the original sample of 54 participants, one did not provide data because of a computer breakdown, three were nonnative speakers who had noticeable comprehension difficulties, and six participants produced too few valid *Yes* responses (typically because they misunderstood the instructions). Reliable data for 44 participants remained (80% female, average age = 19 years).

Materials. Stimulus behaviors were identical to those in Study 3. Three of the probe words differed, however. Desire inferences were elicited by the probe “WANTED?” defined as “Did you detect what the main actor wanted in this situation?” Belief inferences were elicited by the probe “AWAREOF?” defined as “Did you detect what the main actor was aware of (was thinking, saw, heard, knew, etc.) in this situation?” Gender inferences were elicited by the probe “FEMALE?” defined as “Was the main actor FEMALE?”

Instructions for the specific queries following a *Yes* response were adapted as well. For desire inferences, “your specific answer would be what the actor wanted in this situation (e.g., “She wanted to feed her family”).” For belief inferences, “your specific answer would be what the person was aware of, saw, was thinking, etc. (e.g., “He was aware of the car in front” or “thought that the car was expensive”).” The actual queries on the screen after a *Yes* response were “What did the actor want?” and “What was the actor aware of, saw, was thinking, etc.?”

Exclusions, corrections, replacements. We excluded 10 reaction times because they fell outside the within-subject boundaries of ± 2.5 SD, and analysis of audio files suggested 24 eliminations and 6 corrections (< 2% of 1,584 trials). To retain participants in the within-subject analyses we entered 52 mean-replaced values into the matrix of 528 scores.

Mismatched inferences. Only 2% of verbal responses to follow-up queries were inconsistent with the probed inference (e.g., a desire description for an AWAREOF? probe), substantially lower than in the previous studies. Excluding these trials did not alter the results.

Results

Preliminaries. The grand mean for inference likelihood was 65%, slightly lower than the 72% in the video-based dataset of Study 3. The grand mean for speed was 1,581 ms, considerably slower than the 1,395 ms in Study 3, though still faster than the grand means in text-based Study 1 (1,678 ms) and Study 2 (1,674 ms). Gender inferences with the new probe FEMALE? were similarly fast ($M = 1,223$ ms) as inferences with ISMALE? in Study 3 ($M = 1,267$ ms).

Overall inference pattern. (All t values, $df = 43$.) The stable ordering of inferences was replicated. Intentionality inferences ($M = 80%$) were slightly more likely than desire inferences ($M = 75%$, $t = 1.9$, $p = .06$) but equally fast ($M_s = 1,436$ ms and 1,471 ms, respectively, $t < 1$). Belief inferences were less likely ($M = 54%$, $t = 6.1$, $p < .001$) and slower ($M = 1,695$ ms, $t = 7.2$, $p < .001$). Personality inferences were both less likely than the rest ($M = 51%$, $t = 4.6$, $p < .01$) and slower than the rest ($M = 1,721$ ms, $t = 4.9$, $p < .001$), but indistinguishable from belief inferences by themselves. The linear effect explained 34% of the total variance in likelihood (4% nonlinear variance) and 33% of the total variance in speed (3% nonlinear variance).

Independent of this hierarchical inference pattern, the two probe word changes did have some effects. AWAREOF elicited 18 percentage points fewer belief inferences than THINKING in Study 3, and WANTED elicited 14 percentage points fewer desire inferences than GOAL. AWAREOF also slowed down belief inferences by 277 ms, more than WANTED slowed down desire inferences (167 ms), and more than the study as a whole slowed down the two unchanged probes, INTENTIONAL and PERSONALITY (150 ms). This pattern made belief inferences resemble personality inferences more than in the previous studies.

Likelihood of inferences within behavior classes. As before, in response to both untailored and goal-tailored behaviors, intentionality and desire inferences were of similar likelihood ($M_s = 77%$ to 85%, $ps > .11$); belief inferences ($M = 49%$ to 52%) were less likely than the two ($ts > 5.5$, $ps < .001$); and personality inferences ($M = 39%$ to 41%) were least likely of all ($ts > 5.3$, $ps < .001$). For goal-tailored behaviors, personality inferences were also less likely than belief inferences by themselves ($t = 2.2$, $p < .05$).

In response to trait-tailored behaviors, the familiar compression of inference likelihoods emerged. Desire inferences ($M = 63%$) became less likely than intentionality inferences ($M = 76%$, $t = 2.8$, $p < .01$), likelihood of belief inferences was barely lower ($M = 60%$, $t = 1.7$, $p = .10$), and likelihood of personality inferences ($M = 73%$) no longer differed significantly from the average of the remaining inferences ($p > .19$).

Speed of inferences within behavior classes. The speed patterns for untailored and goal-tailored behaviors were identical (see Figure 8). Intentionality and desire inferences were fastest ($M_s = 1,414$ to 1,457 ms) and did not differ ($ts < 1$), belief inferences ($M_s = 1,688$ to 1,694 ms) were slower than the two ($ts > 4.4$, $ps < .001$), and personality inferences were slower than the average of the rest ($M_s = 1,759$ to 1,811 ms, $ts > 4.0$, $ps < .001$). For goal-tailored behaviors, personality inferences were also slower than belief inferences by themselves ($t = 2.2$, $p < .05$).

In response to trait-tailored behaviors, desire inferences ($M = 1,565$ ms) tended to slow down somewhat relative to intentionality inferences ($M = 1,439$ ms, $t = 1.5$); belief inferences ($M = 1,704$ ms) remained slower than the two ($t = 3.1$, $p < .01$); personality inferences sped up as in other studies ($M = 1,592$ ms) and did not differ from the average of the remaining inferences.

⁹ We decided against using the simple word *aware* because the cue might have triggered the question of whether the agent was aware in the sense of conscious.

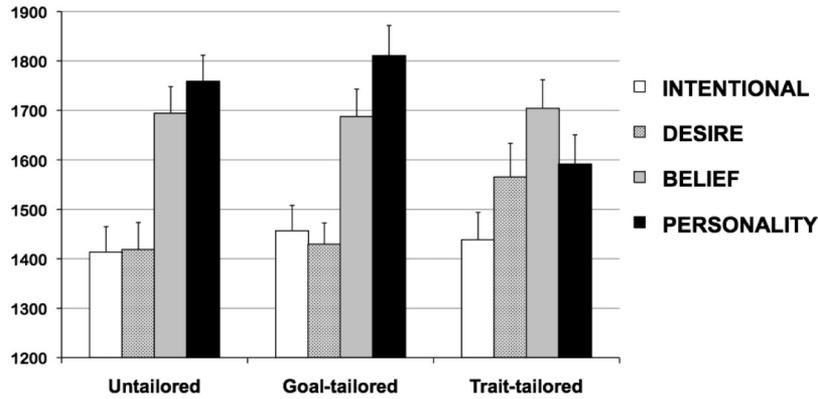


Figure 8. Speed of inferring intentionality, desire (probed with WANTED), belief (probed with AWAREOF), and personality in Study 4 (video stimuli). Error bars indicate standard error of the means.

Specific tailoring effects. Goal-tailored behaviors had no reliable tailoring effect on desire inferences. Compared with untailored behaviors, trait-tailored behaviors had a facilitating effect on personality inferences, both in likelihood (+33%, $t = 6.2$, $p < .001$) and speed (−167 ms, $t = 2.3$, $p < .05$). But as in Studies 1 and 2, trait-tailored behaviors also had an inhibiting effect on desire inferences, making them less likely (−14%, $t = 2.8$, $p < .01$), and slower (+147 ms, $t = 2.1$, $p < .05$) than in response to untailored behaviors.

Discussion

Study 4, using video stimuli, replicated the pattern of inferences found in Studies 1 to 3, even with new probe words for desire and belief inferences. The results further strengthen the evidence for a hierarchy of social inferences, now robust across stimulus formats (text and video) and a variety of eliciting probes. This robustness rules out verbal demands (from stimuli or probes) as an account of the hierarchy inferences and points to genuine process differences between inference types.

Probe word effects. For both desire and belief inferences, the alternate probe words in Study 4 reduced inference likelihood, and for belief inferences, the alternate slowed down access speed. The sole effect on the hierarchical pattern, however, was that in response to untailored and goal-tailored behaviors, belief inferences became indistinguishable from personality inferences (whereas they were usually distinct in the first three studies). Thus, the alternate probe word AWAREOF (like BELIEF in Holbrook, 2006) makes belief inferences even more challenging. As a result, we can be confident that the lower likelihood of belief inferences (relative to desire and intentionality inferences) in Studies 1 to 3 was not the unique result of the THINKING probe; in fact, this probe offers the most efficient way of eliciting belief inferences, which are, however, still less likely and slower than desire and intentionality inferences.

Interestingly, intentionality inferences no longer showed the slowing in response to goal-tailored behaviors that we had observed in Studies 1–3. Because the intentionality probe in Study 4 was unchanged, this newly gained stability of intentionality inferences may have resulted from a context effect: Perhaps the greater challenges posed by the alternate probes for desire and belief

inferences made intentionality inferences seem, by comparison, more straightforward, even for the more multi-layered behaviors in the goal-tailored set.

Overall, then, the hierarchy of social inferences does not rely on specific probe words. Across the studies by Holbrook (2006) and the present studies, we used two words for intentionality (INTENTIONAL, INTENT), three for desire (GOAL, THEGOAL, WANTED), three for belief (BELIEF, THINKING, AWAREOF), and two for personality (PERSONALITY, TRAIT). Despite variations in probe averages, the hierarchical pattern consistently held up.

Average speed differences. In video-based Study 3, average access speed was 281 ms faster than in text-based Studies 1 and 2, whereas Study 4 was only 95 ms faster. Why? The alternate belief and desire probes in Study 4 slowed down inferences relative to Study 3 by an average of 222 ms. However, the standard probes for intentionality and personality also slowed down between Studies 3 and 4, by an average of 150 ms. Further exploring differences between studies, we discovered an additional source of overall speed variation: the period within the academic term of conducting the experiments. The majority of Study 3 participants were run in the very first week of fall term (Oct 2–4) and averaged 1,381 ms, whereas a subset was run October 9–12 and averaged 1,459 ms, reflecting a term-period effect of 78 ms. In Study 4, the October 9–12 period (the earliest in which participants were run) was faster than any other period of the term, yielding a term-period effect of 81 ms. Similarly, for a comparison of October 9 and the rest of the term, the two text studies showed a term slowing of 85 ms (see Table 3). Most important, these term-period effects held the same way for all four inference types. Whatever factors explain the effects—motivation, stress, personality differences—they are orthogonal to the hierarchical pattern of social inferences.

We can now explain why Study 4 was slower than Study 3. First, in Study 4 nobody was run during the fastest period of October 2–4; second, Study 4 contained two alternate inference probes. Furthermore, holding constant time of term (October 9–12), we can compute a more precise estimate of the probe word effects: Participants responded 133 ms more slowly to WANTED than to THEGOAL and 216 ms more slowly to AWAREOF than to THINKING. Likewise, holding constant both time of term and inference probes, we can compute a more precise estimate of the

Table 3
Term Period Effects and Text Versus Video Effects on Average Speed in Studies 1–4

	Overall	First week (Oct 2-4)	Early (Oct 9-12)	Rest of Term
		Video speeding		Term slowing
Studies 1-2 (Text)	1676 ms		-141 ms 1600 ms (N = 9)	1685 ms (N = 74) 85 ms
Study 3 (Video)	1395 ms	1381 ms (N = 33)	1459 ms (N = 7)	78 ms
Study 4 (Video)	1581 ms		1537 ms (N = 20)	1618 ms (N = 24) 81 ms

Note. Term slowing refers to the slowing of reaction times after the first two weeks of the academic term. Video speeding refers to faster reaction times for video stimuli compared with text stimuli. The results of these comparisons are indicated in bold, computed from values enclosed in the same rectangular frame.

video speeding effect: video stimuli in Study 3 elicited 141 ms faster responses than text stimuli in Studies 1 and 2 (see Table 3).

We tested these explorations more systematically in a multivariate analysis of covariance at the trial level, with behavior class and inference type as the independent variables, access speed as the dependent variable, and the abovementioned covariates of stimulus presentation (text/video), probe variant (standard/alterd), and time of assessment (early/late in the term). In this sensitive test on over 4,000 inferences, all three covariates showed independent prediction of overall access speed ($t_s > 3.5$, $p_s < .001$), but the familiar speed patterns within and across behavior classes were unaltered. To wit, when we examined the nine critical contrasts (three hierarchical inference type comparisons within each of the three behavior classes) before and after correcting for covariates, the corresponding t value differences ranged from -0.21 to 0.45 , with an average of 0.09 .

Across four studies. Summarizing the results gathered so far, Figure 9 shows the results for the critical variable of access speed aggregated across Studies 1 through 4. We see the hierarchical patterns most clearly within untailored behaviors and within the entire sample of 36 behaviors, and we also see the specific deviations (indicated with up- and down-arrows) that goal-tailored and trait-tailored behaviors caused relative to untailored behaviors. The consistency of at least a partial hierarchy across all the 36 behaviors is noteworthy. If we express the hierarchy as a normalized linear contrast (Intentionality $\times -0.671$ + Desire $\times -0.224$ + Belief $\times 0.224$ + Personality $\times 0.671$) and compute it for each of the 36 behaviors, we find linear effects ranging from -144 to 738 with an average of 201 (given an overall reaction time average of $1,580$ ms). Out of the 36 behaviors, 31 show a positive linear effect. Even in trait-tailored behaviors, the linear effect's average is 118 and 10 out of 12 behaviors show positive effects. Finally,

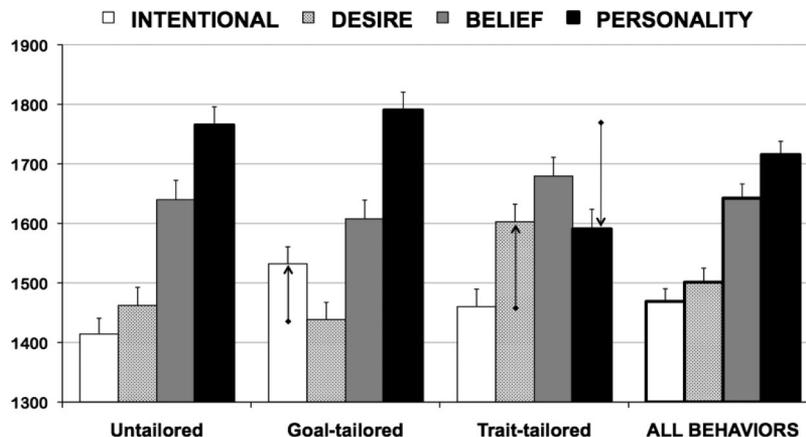


Figure 9. Social inference hierarchy for speed of making four types of inference within and across three classes of behavior (unweighted means across two text-based and two video-based studies). Error bars indicate within-cell errors. Arrows show effects of tailored behavior types relative to untailored behavior types.

personality inferences are accessed faster than either intentionality or desire inferences in only 6 out of 36 behaviors.

Study 5: Online Inferences

The first four studies assessed a social perceiver's ability to make inferences from verbal or visual behavior and, if the inference was made, measured the perceiver's speed of accessing the inference right after the behavior observation. This speed indicates the availability of the inferred information for further processing, which is important when mental state information is needed to make a subsequent judgment (such as blame; Guglielmo & Malle, 2011), when it guides actions towards the other person, and when it is verbally reported.

We have assumed that participants who affirm having made an inference had formed that inference *while* processing the stimulus. However, one might assume instead that participants encode and store the stimulus information as a whole and, when probed for a particular inference, derive that inference on the spot from the stored stimulus information. Some of our data make such on-the-spot construction less plausible. For one thing, information about gender does not have to be constructed but is simply retrieved, and the access speed of intentionality and goal inferences was only slightly slower than the speed of gender inferences in Studies 3 and 4 and faster in Studies 1 and 2. There is also independent evidence that goal inferences and personality inferences emerge from online processing of text stimuli (Graesser, Singer, & Trabasso, 1994; Hassin et al., 2005; Uleman et al., 2008; Van der Cruyssen et al., 2009; Van Duynslaeger et al., 2007). Further, the range of access times in our studies (1,200–1,600 ms) is the same as that of a parallel but much simpler task in the text processing literature, in which participants read a text and immediately afterwards verify that a probe word was indeed mentioned in the text (Dopkins, Klin, & Myers, 1993). To solve this task participants must match the probe word against a retrieved memory trace of the previously encoded word in the text. Given that it takes arguably no more time to retrieve a more complex desire or belief inference than to retrieve a plain word, it is reasonable to assume that such an inference has already been made during stimulus processing. Nonetheless, we cannot entirely rule out that people may construct an inference even within this short time frame or that a substantial number of inferences are constructed in retrospect (e.g., the slower ones). We therefore directly tested the online processing hypothesis in Study 5.

The Online Inference Paradigm

If we can capture the time points at which people have accumulated sufficient evidence to make a particular inference and if these accumulation times show the hierarchy discovered in the previous studies, we would gain strong evidence for the online interpretation of those studies and for the generality of the inference hierarchy. To make this goal experimentally tractable, we used video stimuli, which depict behaviors unfolding in time and therefore allow for continuous online information accumulation. Furthermore, before watching each video, participants were instructed to look for information supporting one of five inferences, which we called the "target" (e.g., "Look for the goal"). Participants had to stop the video as soon as they felt reasonably confi-

dent that they were able to infer that target. This *stopping point* defined the time at which the online processing had culminated in the target inference, without retrieval or experimenter probing. A subsequent follow-up query (e.g., "What goal did you detect?") ensured that people indeed had made an appropriate inference at the time they stopped the video.

Thus, Study 5 put the potential inference hierarchy to a stringent test. If intentionality and desire inferences are not only faster to access but are actually formed more quickly than belief inferences, we should see shorter stopping times for desire and intentionality than for belief inferences. Moreover, if mental state inferences, particularly desire and belief inferences, precede and are sometimes prerequisites for making personality trait inferences, then we should see the longest stopping times for such trait inferences.

Method

Participants. Undergraduate students who were enrolled in introductory psychology or linguistics courses participated in exchange for partial course credit. Of the original sample of 75 participants, one did not provide data because of a computer breakdown, two were non-native speakers who had noticeable comprehension difficulties, and five participants produced fewer than 33% valid *Yes* responses (typically because they misunderstood the instructions). Reliable data for 67 participants remained (80% female, mean age = 19 years).

Material. The stimuli were the same video clips as in Studies 3 and 4. This time six videos were reserved for the gender control question, namely the three gender trial videos in the previous studies and the three "Do not respond" catch trials. Two video clips of unintentional behaviors and eight practice trials also remained from the previous studies.

The natural arc of unfolding information can be expected to differ from one video to the next, so we paired each of the 36 behaviors with each critical probe an equal number of times across participants. We also pseudorandomized the order of trials with the constraint of no more than two consecutive trials with the same inference probe.

Procedure. The task instructions were as follows:

Each video behavior contains a variety of information. Before each video we will tell you what bit of information you should look for (your "target"), and as soon as you find it, you stop the video.

Participants learned that the targets would be identified by single words on the screen just before each video commenced, embedded in the phrase "Stop when you notice [word]." The four critical probe words were INTENTIONAL, THEGOAL, THINKING, and PERSONALITY; the control probe was GENDER. Each of these probes stood for a more complete question introduced in full as part of the instruction (e.g., PERSONALITY = "Stop the video as soon as you notice a PERSONALITY characteristic/trait the main actor has."). Participants were instructed to respond by pressing the space bar *as quickly as possible* and not to wait until they were completely confident in their inference. We encouraged them "to 'beat' the video and stop it before it stops on its own."

After participants stopped a given video and thus indicated they had arrived at a certain inference, the follow-up query appeared and asked participants to say out loud what it was that they found (e.g., "Please say what personality trait you noticed"; "what be-

havior was intentional"). The instructions provided detailed examples for each query. For example, for PERSONALITY, they read,

Let's say your target is the actor's PERSONALITY and you watch a video of a woman taking her kids to the playground. As soon as the video suggests something about that woman's personality you would hit Stop (the space bar), and then you would say *what* personality characteristic you thought you noticed (e.g., "caring").

As in the previous studies, participants first rehearsed the meaning of the probes and then worked through eight practice videos.

Skewness. At the trial level, reaction times showed some skewness, primarily because stopping times centered on 3,900 ms, but the possible values of stopping times ranged from around 400 ms to 13 s (for the longest video). Aggregating reaction times across the three videos within each of the 12 design cells, however, greatly reduced skewness. One cell had two outliers (defined as $>2.5 SD$ above the mean), nine cells had one outlier, and two cells had none. When we recoded these outliers as $M + 2.5 SD$ and reran all analyses, we found identical results. Thus, all reports below are based on the original scores.

Exclusions, corrections, replacements. Analysis of audio files suggested 11 exclusions and four corrections out of 2,412 trials. To retain participants in the within-subject analysis, we mean-replaced 29 missing values in the matrix of 804 critical scores.

Follow-up queries. People's verbal responses to follow-up queries (after they had stopped a video) showed mismatches in 7.7% of trials. Desire inferences were again the most common ones to intrude (4.9% of trials), and intentionality inferences were most susceptible to such desire intrusions (12.4% of trials). Excluding such mismatched items did not alter the results.

Results

Stopping rates no longer represent the likelihood of spontaneous inferences but the ability to make, when requested, a specific inference before the end of the video. Overall, participants stopped 77% of the videos. Stopping rates for intentionality, desire, and belief inferences were statistically indistinguishable (ranging from 74% to 84% within behavior classes), but each was higher than the

rates for personality inferences (62% to 73%; $ts(66) > 3.0$, $p < .001$).

Online inference speed. (For all t values, $df = 66$.) Across all behaviors, intentionality inferences ($M = 3,528$ ms) and desire inferences ($M = 3,603$ ms) were similarly fast ($t < 1$), belief inferences were slower ($M = 3,918$ ms, $t = 3.6$, $p < .001$), and personality inferences ($M = 4,600$ ms) were slowest of all ($t = 7.5$, $p < .001$), even slower than belief inferences by themselves ($p < .001$). The linear effect explained 26% of the variance (with 4% nonlinear explained variance), and the pattern was highly consistent across behaviors (75% had a positive linear effect term) and across participants (87% had a positive linear effect term).

As Figure 10 shows, in response to untailored behaviors, intentionality inferences ($M = 3,296$ ms) and desire inferences ($M = 3,080$ ms) were fastest and did not differ reliably ($t < 1.4$). Belief inferences ($M = 3,630$ ms) were slower than the two ($t = 3.4$, $p < .01$), and personality inferences were slowest of all ($M = 4,502$ ms, $t = 7.1$, $p < .001$).

In response to goal-tailored behaviors, intentionality inferences ($M = 3,561$ ms) and desire inferences ($M = 3,761$ ms) were again similarly fast ($t < 1$), and belief inferences ($M = 4,113$ ms) were slower than the two ($t = 2.8$, $p < .01$). Personality inferences ($M = 4,338$ ms) were slower than the average of the rest ($t = 2.6$, $p = .01$) but not distinguishable from belief inferences ($t < 1$).

In response to trait-tailored behaviors, intentionality inferences ($M = 3,728$ ms), desire inferences ($M = 3,968$ ms), and belief inferences ($M = 4,009$ ms) were equally fast ($ts < 1$), and personality inferences ($M = 4,960$ ms) were slower than each and all of them ($t = 6.3$, $p < .001$).

Tailoring effects. Desire inferences did not speed up in response to goal-tailored behaviors nor did personality inferences speed up in response to trait-tailored behaviors. Merely tailoring a behavior towards a certain inference apparently does not ensure that the diagnostic information comes early in the unfolding event structure. On the contrary, tailoring may involve a careful (and more extensive) setup to draw out a certain inference, and such a setup demands additional processing time for *all* inferences. Indeed, for goal-tailored behaviors people stopped videos, on average, 316 ms later than for untailored behaviors, and for trait-tailored behaviors, they stopped them 539 ms later. The inference

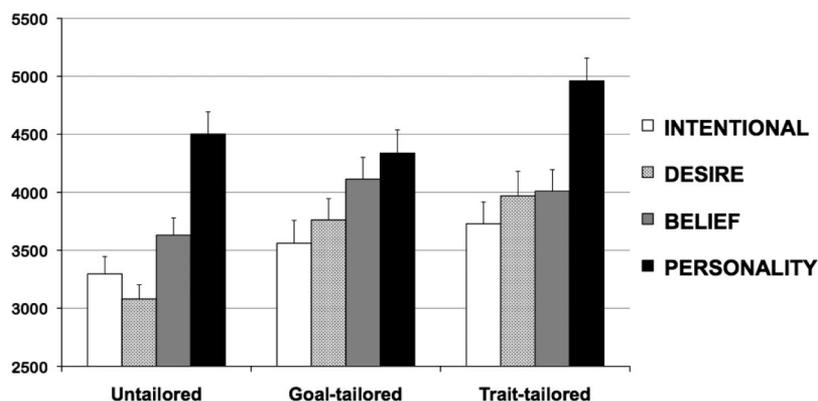


Figure 10. Stopping times (in ms) for four types of inference within three classes of behavior in Study 5. Error bars indicate within-cell standard errors.

pattern across all 36 behaviors, however, adjusts for those effects and displays a clear hierarchy of social inferences.

Discussion

People stopped 73% of the videos they watched, and when they did, on average half of the video had elapsed ($M = 3,876$ ms, $SD = 915$ ms). The fact that people were willing and able to make these online inferences for so many videos and with such speed supports the suggested interpretation of access speeds in Studies 1–4: What people retrieved within about 1,500 ms were inferences made during the prior online processing of the stimulus behaviors.

Equally important is the order in which the four major inferences were formed during online processing. Replicating the earlier pattern of access speeds, we found that people made online intentionality and desire inferences faster than belief inferences and belief inferences faster than personality inferences. Thus, a hierarchy of social inferences holds both in the initial formation of inferences and in the availability of those inferences for immediate access. Such parallel ordering across encoding, storage, and retrieval narrows the possible theoretical models that can account for the data. We will address this question in the concluding discussion.

One question of interpretation of Studies 1–4 remains. Is it possible that people made hierarchical online inferences in Study 5 (when they were encouraged to do so), but not in Studies 1–4? We already mentioned two signs that weaken this hypothesis, namely the range of access times—which were no slower for inferences of intentionality or goal than for gender—and the absolute level of access times—comparable with the retrieval of a word meaning. But might people have quickly, after being probed, laid claim to the inference and then taken their time to construct a fitting content for the verbal follow-up response? This would not explain the systematic differences in reaction times among inference types because such inference claims could be made equally fast for any inference. Moreover, this hypothesis is contradicted by the pattern of verbal latencies. An analysis on a subset of participants in text-based Study 2 showed that 52% of verbal follow-ups had latencies between 1 and 2 s, and 97% were below 4 s. Similarly, an analysis of a sample of valid trials in video-based Study 3 showed an average response latency of 1.56 s. We cannot rule out that some people, for some of the trials, constructed inference content post hoc, but such trials would, if anything, dilute the differences among inference types that we actually found.

General Discussion

The important role of mental state inferences in human social cognition is unquestioned, but relatively little is known about the cognitive processes that underlie such inferences. The present studies take one step toward advancing our knowledge about these processes. In five experiments, we examined people's simultaneous inferences about intentionality, desires, beliefs, and personality. Stimulus behaviors were presented both as verbal descriptions and as videos. Inferences were assessed both in a retrieval paradigm (measuring the likelihood of inferences and the speed of accessing them immediately after they were made) and in an online processing paradigm (measuring the speed of forming in-

ferences during behavior observation). The results provide consistent evidence for the following conclusions:

1. Intentionality and desire inferences are primary, both in likelihood and in speed of inference.
2. Compared to these two, inferences about an agent's beliefs are less likely and slower.
3. Inferences of personality traits are least likely and slowest among all tested inferences, except when the stimuli were specifically tailored to elicit such inferences.

We now discuss in more detail each of these conclusions.

Intentionality and Desires

Intentionality and desire inferences appear to be just as basic in the social perception process as they are in the early development of social perception. Researchers have found that 6-month-old infants recognize the goal-directedness of actions (Woodward, 1998), which requires detecting an intentional behavior and tracking systematic connections between an agent and a goal object. In our studies, the goals people tracked were usually not objects but actions or states, and in many cases, they were not observable but implied or anticipated; this makes the formation and access speed of those inferences all the more impressive.

Even though desires and intentionality were comparable in how likely and how quickly they were inferred, the two types of inferences are not interchangeable. Intentionality inferences were inhibited by goal-tailored behaviors, whereas desire inferences were inhibited by trait-tailored behaviors; desire inferences occasionally intruded into belief inferences, whereas intentionality inferences did not; and intentionality inferences were consistently faster (albeit to a small degree) for untailored and trait-tailored behaviors.

We suspect that intentionality inferences will be even faster than desire inferences for target behaviors that are expressed in basic movements, that are unitary (removing competition over which of several candidates is the primary intentional act), and that are unusual (leaving the person's goal opaque). As an illustration for the latter case, the unusual action of a plumber slipping \$50 into his wife's purse averaged a retrieval time of 1,094 ms for intentionality inferences but 1,778 ms for desire inferences.

Beliefs and Desires

Philosophy, developmental research, and more recently, social psychological research have documented the distinct conceptual, psychological, and linguistic properties of beliefs and desires as two central but distinct targets of social cognition (Dretske, 1988; Malle, 2004; Wellman & Woolley, 1990). Understanding desires emerges earlier in development than understanding beliefs (Wellman et al., 2001); children's explanations of behavior cite desires earlier than beliefs (Bartsch & Wellman, 1995); adults have a more difficult time providing belief reasons than desire reasons when explaining behavior (Malle et al., 2007); and inferences of another person's beliefs appear to require effortful correction of one's own beliefs (Apperly et al., 2008; Barr & Keysar, 2005; Birch & Bloom, 2007; Epley, Morewedge, & Keysar, 2004).

Our results add to these findings by characterizing the belief–desire distinction at the process level: People arrive at desire inferences earlier than at belief inferences in their online processing of behavioral information; they spontaneously infer desires more often than beliefs; and they access desire inferences, once made, more quickly than belief inferences.

Why would belief inferences take longer than desire inferences? Desires are often revealed in observable behavior (Dik & Aarts, 2007; Malle, 2005) and are constrained by the norms and demands of context (Bruner, 1990), whereas beliefs are more purely “in the head” and can be highly idiosyncratic or even unrelated to the observed context. If this is correct, then eye-tracking methodology should demonstrate that desire inferences are even more efficiently inferred from the perceiver’s monitoring of specific behaviors, expressions, and agent–context interactions.

Personality

Two competing hypotheses existed about personality trait inferences. According to the first, dominant in social psychology, “the attribution of personality traits to other people is ubiquitous” (Uleman, 2005, p. 253), and “traits are often attributed in a very rapid, spontaneous, almost inevitable fashion” (Fiedler & Schenck, 2001, p. 1533). According to the second hypothesis, trait inferences do not have priority because they emerge later in development and possibly evolved later than attributions of intentionality and mental states (Kalish & Shiverick, 2004; Snodgrass, 1976). The data from our studies favor a process variant of the latter hypothesis. People can and do make trait inferences, especially when prompted by the affordances of certain behaviors, but they do not make them inevitably or even very frequently. When they do make personality inferences, they are more slowly formed and retrieved than inferences about intentionality and mental states. For cases in which people do not make a trait inference from a given behavior, they still infer that behavior’s intentionality, the agent’s desires, and somewhat less often, the agent’s beliefs.

This pattern of less frequent and slower personality inferences is consistent with data on the lower priority of personality traits, both as objects of attention in social interaction (Malle & Pearce, 2001) and as contents that make up explanations of everyday behavior (Malle et al., 2007). The pattern is also consistent with models that describe trait inferences as relying on prior goal inferences (Read et al., 1990; Reeder, 2009) and with evidence showing that across age levels in development, children’s success in mental state inference tests predicts their success in personality inference tests, but not the other way around (Ramsay, 2003).

The reluctance to make personality inferences that we found in the current paradigm seemingly contradicts people’s claimed frequency and readiness to make such inferences from single behaviors (Gilbert & Malone, 1995; Ross, 1977). However, all the evidence for the frequency of personality inferences comes from tightly constrained lab experiments in which people are asked solely to infer traits from stimuli that are often tailored to evoke such concepts. These stimuli correspond to our trait-tailored behaviors, which increase, unsurprisingly, people’s tendency to make personality inferences. In naturalistic settings, the prevalence of trait inferences appears to be far lower (De Raad, 1984; Lewis, 1995; Malle et al., 2007), mirroring our findings with untailored behaviors.

In natural contexts, people are more likely to infer personality if a decent amount of behavioral information has accumulated—for example, information of frequent acts (Buss & Craik, 1983) or conditional act–situation patterns (Wright & Mischel, 1987). Although all inference types require information accumulation, personality inferences require more of it. As a result, when considering single behaviors (as sentences or short videos), people do not willy-nilly infer traits, unless they are encouraged to. Even when invited to make such personality inferences (in Study 5), people declined to do so on almost 40% of the general trials (involving untailored and goal-tailored behaviors) and still on 30% of the trait-tailored trials.

How do we reconcile the present conclusions with the consistent evidence for spontaneous trait inferences (STI; Skowronski, Carlston, & Hartnett, 2008; Uleman, Saribay, & Gonzalez, 2008)? To allow a comparison between the different methodological paradigms, we had created an overlap in stimulus behaviors—the trait-tailored behaviors used in STI research. For these behaviors, our studies show that people make trait inferences in about 70% of cases, which is the rate that pretesting of those behaviors for STI studies had established. This rate did not, however, exceed the rate of intentionality, desire, and belief inferences, even though the stimuli were not tailored to those inferences. Conversely, for stimulus behaviors that were not specifically tailored to infer personality, the trait inference rate dropped precipitously. Moreover, in all assessments of formation and access speed, trait inferences were no faster and, most often, considerably slower than other inferences. By comparison, STI studies have shown that in response to trait-tailored stimuli, people are capable of implicitly encoding trait concepts. What we do not know is how many other implicit inferences people formed that were not probed (e.g., about intentionality or beliefs) and whether implicit trait inferences extend to stimuli that are not carefully tailored. All in all, our findings do not contradict STI studies, but they challenge STI researchers to extend both stimulus behaviors and inference types to assess personality trait inferences in a broadly generalizable context.

Limitations

Our studies share with previous work a focus on inferences from intentional behaviors. This focus makes good sense given the social perceiver’s greater attention to intentional behavior (Malle & Pearce, 2001) and given that inferences of intentionality and goals obviously must be studied with intentional behaviors. Personality trait inferences, however, could be examined in other ways as well. Despite Jones and Davis’s (1965) proposal that people confidently infer traits only from intentional behaviors, it seems clear that some unintentional behaviors (e.g., shaky voice) can elicit trait inferences (e.g., about nervousness). These traits may not be the only inferences made in such cases. Unintentional behaviors are often undesired side effects of other intentional behaviors, so people may infer desires or beliefs that the agent holds while, or in advance of, performing the unintentional behavior. In fact, someone’s shaky voice reveals nervousness only if we assume that the person, say, *wants to give a competent talk*, whereas a shaky voice in a relaxed conversation may imply an illness or speech defect. The upshot is that the hierarchy proposed here may still be largely intact even with unintentional behaviors, beginning with the inference that the behavior is *not* intentional, inferences about what the person *would like* to do or *thought would*

happen, and inferences about the traits revealed by the unintended behavior or outcome. Future research will need to test the onset and access times of these multiple inferences.

A related limitation is the focus on social inferences from *behavior*. Again, the focus makes sense given that the challenge of social cognition arises primarily in social interaction—both in adult life, the infant’s early steps toward social cognition, and in the most plausible scenarios of evolutionary origins of social cognition. But research suggests that when instructed to do so, people draw personality inferences from 2-s segments of behavior (Ambady & Rosenthal, 1993) and from nonbehavioral information such as photographed faces (Willis & Todorov, 2006). However, people also make mental state inferences from such stimuli (Ambady, Bernieri, & Richeson, 2000), and mental state inferences may even underlie these personality inferences (Todorov, 2008; Trope, 1986). New research that directly compares both onset and access times among these multiple inferences from minimal behavioral stimuli could generalize the notion of a hierarchy of inferences or else show its boundary conditions.

A final limitation to consider is the sampling of stimulus behaviors. For comparisons with previous research, we used trait- and goal-tailored behaviors and added a first set of untailored stimuli. But there is little theory and there are few methodological principles that guide the selection of stimulus behaviors in social perception research (Ambady et al., 2000), especially when aiming at representative design (Brunswick, 1956). One possible future approach would document what people do in everyday life (Mehl & Pennebaker, 2003) and build a proportional behavior set for use in experiments. Another would emulate the lexicographic analysis of personality (Goldberg, 1992) and base stimulus selections on the usage frequency of behavior verbs.

Toward a Theory of Social Inferences

Even though our data are compatible with a number of different explanatory accounts, the present studies provide at least three novel facts that should guide a theory of social inferences.

The first fact is that the process hierarchy parallels the hierarchy of concepts found in developmental and comparative research, so a critical determinant of an inference’s position in the hierarchy appears to be concept complexity. Concept complexity may be decomposed into a degree of behavior expression (e.g., goals express themselves strongly; fantasies, weakly; Malle, 2005) and semantic precision (detecting “creative” would be more difficult to infer than “punctual”; Uleman, 2005). Thus, concepts that have clear behavior expression (e.g., ethnicity) or are relatively precise (e.g., self-propelledness) should be likely and fast to infer, whereas concepts that have little to no behavior expression (e.g., mental images) or have fuzzy boundaries (e.g., resentment) should be less likely and slower to infer. Such differences should also create variation within inference types, such that, for example, trait observability (Funder & Dobroth, 1987) would speed up trait inferences. At the process level, behavior expression should primarily affect inference formation, whereas precision should primarily affect retrieval (because even in light of relevant evidence, the right label must be found).

The second fact is that the hierarchy is modality independent—verbal and visual stimuli elicit the same ordering of inferences. The underlying inference mechanism must therefore be capable of

integrating visual feature analysis, semantic interpretation, and general top-down knowledge such as action scripts and canonical object uses (Zacks, Speer, Swallow, Braver, & Reynolds, 2007). This makes single-process accounts of social inference less plausible, such as one based on the “mirror system” (Gallese, Keysers, & Rizzolatti, 2004). Instead, a variety of processes seem to be involved in generating social perception, such as projection, simulation, perspective taking, and knowledge structures (Malle, 2004, 2005, 2008).

The third fact is that the hierarchy exists in the speed of both forming each inference and accessing that inference. This finding suggests that inferences formed earlier in the observation process remain in the “forefront” of memory. This is not obvious because later-formed inferences could benefit from a recency effect and, therefore, be accessed more quickly. The fast encoding and access of certain primary inferences (e.g., intentionality and desire) also opens the possibility that these primary inferences may be available for constructive retrieval of secondary inferences that were not fully formed during behavior observation. For example, personality inferences may not always be made online, but if the available behavior information was already processed with respect to its intentionality and goal, this inferential content is directly accessible and can be used to construct a trait inference (Read & Miller, 2005; Reeder, 2009). This potential for secondary construction further encourages us to revisit the literature on STI and ask whether perceivers originally encode mental state information about an agent and, perhaps only later, upon encountering a trait label, assess its compatibility with the previously stored mental state information. Traits would thus not be spontaneously *inferred* but *endorsed*. Only new experiments that integrate multiple inference types can test this hypothesis.

Studying the exact relationships among inference types will provide important additional building blocks for a theory of social inferences. Are the inferences engaged in a race of parallel processing (Freeman & Ambady, 2011), or are there serial priorities? Are there meaningful dependencies among the inferences, such as facilitation, competition, or interference?

Ongoing and Future Research

Stimulus medium. The fact that consistent inference patterns emerged for both text and video stimuli has two complementary implications for future work. On the one hand, we can conclude that verbal descriptions can in principle elicit social inference processes that generalize to the most natural of contexts, namely, behavior observation. On the other hand, we have seen that responses to video stimuli are just as reliable as text stimuli and do not impose overly demanding complexity on the participants. Most important, with the high realism of video stimuli we may detect many psychological patterns that would not be detectable with text descriptions alone, such as differential impact of body language, facial expressions, gaze, or visible stigmas and the temporal dynamic of multiple social inferences as people track online the unfolding behavior stream.

Expansions. We are currently examining additional types of inferences such as emotions and moral judgments (Guglielmo & Malle, 2009), especially the comparison between intentionality and blame (Malle & Guglielmo, 2011). Equally interesting would be the inclusion of more basic social concepts such as age and

ethnicity. Both inference formation and retrieval are relevant here because initial activation of a social category may not always be used in later judgment (Kunda, Davies, Adams, & Spencer, 2002). Moreover, it may be possible to disentangle in this paradigm trait inferences strictly made from behavior and trait inferences derived from a stereotype, perhaps elucidating the long-standing question of how people handle stereotype-inconsistent information. We have also begun to examine inferences about group agents (Dillon & Malle, 2011), and other researchers might contrast inferences about individual group members (Hamilton & Sherman, 1996) with inferences about the group as a whole (O’Laughlin & Malle, 2002). Finally, future work needs to address possible moderators of the proposed hierarchy, especially factors that may favor one type of social inference over another: individual differences and affect on the perceiver side, context and task demands, and culture.

We hope that many other questions will arise within this approach of studying multiple simultaneous social inferences. After all, this is the way people make inferences in the real world: not one at a time, but many at a time, most of the time, and apparently within a systematic hierarchy of processing.

References

- Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 32, pp. 201–271). San Diego, CA: Academic Press.
- Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology, 64*, 431–441. doi:10.1037/0022-3514.64.3.431
- Ames, D. R. (2004). Inside the mind reader’s tool kit: Projection and stereotyping in mental state inference. *Journal of Personality and Social Psychology, 87*, 340–353. doi:10.1037/0022-3514.87.3.340
- Amit, E., Algom, D., & Trope, Y. (2009). Distance-dependent processing of pictures and words. *Journal of Experimental Psychology: General, 138*, 400–415. doi:10.1037/a0015835
- Apperly, I. A., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adults’ performance on a non-inferential theory of mind task. *Cognition, 106*, 1093–1108. doi:10.1016/j.cognition.2007.05.005
- Astington, J. W. (2001). The paradox of intention: Assessing children’s metarepresentational understanding. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 85–103). Cambridge, MA: MIT Press.
- Baldwin, D. A., & Baird, J. A. (2001). Discerning intentions in dynamic human action. *Trends in Cognitive Sciences, 5*, 171–178. doi:10.1016/S1364-6613(00)01615-6
- Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development, 72*, 708–717. doi:10.1111/1467-8624.00310
- Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D. J. (Eds.). (2000). *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed.). New York, NY: Oxford University Press.
- Barr, D. J., & Keysar, B. (2005). Mindreading in an exotic case: The adult human. In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 271–283). New York, NY: Guilford Press.
- Barrett, H. C., Todd, P. M., Miller, G. F., & Blythe, P. W. (2005). Accurate judgments of intention from motion cues alone: A cross-cultural study. *Evolution and Human Behavior, 26*, 313–331. doi:10.1016/j.evolhumbehav.2004.08.015
- Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. New York, NY: Oxford University Press.
- Bassili, J. N. (1993). Procedural efficiency and the spontaneity of trait inference. *Personality and Social Psychology Bulletin, 19*, 200–205. doi:10.1177/0146167293192009
- Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science, 18*, 382–386. doi:10.1111/j.1467-9280.2007.01909.x
- Bogdan, R. (2000). *Minding minds: Evolving a reflexing mind by interpreting others*. Cambridge, MA: MIT Press.
- Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley, CA: University of California Press.
- Buss, D. M., & Craik, K. H. (1983). The act frequency approach to personality. *Psychological Review, 90*, 105–126. doi:10.1037/0033-295X.90.2.105
- Call, J., Hare, B., Carpenter, M., & Tomasello, M. (2004). “Unwilling” versus “unable”: Chimpanzees’ understanding of human intentional action. *Developmental Science, 7*, 488–498. doi:10.1111/j.1467-7687.2004.00368.x
- Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences, 12*, 187–192. doi:10.1016/j.tics.2008.02.010
- Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *The Journal of Political Economy, 97*, 1232–1254. doi:10.1086/261651
- Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior & Development, 21*, 315–330. doi:10.1016/S0163-6383(98)90009-1
- Chandler, M. J., Greenspan, S., & Barenboim, C. (1973). Judgments of intentionality in response to videotaped and verbally presented moral dilemmas: The medium is the message. *Child Development, 44*, 315–320. doi:10.2307/1128053
- Chen, S. (2003). Psychological-state theories about significant others: Implications for the content and structure of significant-other representations. *Personality and Social Psychology Bulletin, 29*, 1285–1302. doi:10.1177/0146167203255226
- Decety, J., Michalska, K. J., & Kinzler, K. D. (2012). The contribution of emotion and cognition to moral sensitivity: A neurodevelopmental study. *Cerebral Cortex, 22*, 209–220.
- De Raad, B. (1984). Person talk in everyday situations. In H. Bonarius, G. Van Heck, & N. Smid (Eds.), *Personality psychology in Europe: Theoretical and empirical developments* (pp. 31–44). Lisse, the Netherlands: Sweits & Zeitlinger.
- Dik, G., & Aarts, H. (2007). Behavioral cues to others’ motivation and goal pursuits: The perception of effort facilitates goal inference and contagion. *Journal of Experimental Social Psychology, 43*, 727–737. doi:10.1016/j.jesp.2006.09.002
- Dillon, K. D., & Malle, B. F. (2011). *A robust hierarchy of social inferences about individuals and group agents*. Paper presented at the annual meeting of the Society for Philosophy and Psychology, Montréal, Quebec, Canada.
- Dopkins, S., Klin, C., & Myers, J. L. (1993). Accessibility of information about goals during the processing of narrative texts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 70–80. doi:10.1037/0278-7393.19.1.70
- Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.
- Ekman, P. (1982). *Emotion in the human face* (2nd ed.). Cambridge, England: Cambridge University Press.
- Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction.

- Journal of Experimental Social Psychology*, 40, 760–768. doi:10.1016/j.jesp.2004.02.002
- Fiedler, K., & Schenck, W. (2001). Spontaneous inferences from pictorially presented behaviors. *Personality and Social Psychology Bulletin*, 27, 1533–1546. doi:10.1177/01461672012711013
- Fiske, S. T., & Taylor, S. E. (2008). *Social cognition: From brains to culture*. Boston, MA: McGraw-Hill.
- Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, 118, 247–279. doi:10.1037/a0022327
- Frith, C. D., & Corcoran, R. (1996). Exploring “theory of mind” in people with schizophrenia. *Psychological Medicine*, 26, 521–530. doi:10.1017/S0033291700035601
- Funder, D. C. (1991). Global traits: A neo-Allportian approach to personality. *Psychological Science*, 2, 31–39. doi:10.1111/j.1467-9280.1991.tb00093.x
- Funder, D. C., & Drobth, K. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology*, 52, 409–418. doi:10.1037/0022-3514.52.2.409
- Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, 8, 396–403. doi:10.1016/j.tics.2004.07.002
- Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought: Limits of awareness, intention, and control* (pp. 189–211). New York, NY: Guilford Press.
- Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 89–150). New York, NY: McGraw-Hill.
- Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, 117, 21–38. doi:10.1037/0033-2909.117.1.21
- Golan, O., Baron-Cohen, S., Hill, J. J., & Golan, Y. (2006). The “Reading the Mind in Films” task: Complex emotion recognition in adults with and without autism spectrum conditions. *Social Neuroscience*, 1, 111–123. doi:10.1080/17470910600980986
- Goldberg, L. R. (1992). The development of markers for the Big-Five Factor structure. *Psychological Assessment*, 4, 26–42. doi:10.1037/1040-3590.4.1.26
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371–395. doi:10.1037/0033-295X.101.3.371
- Guglielmo, S., & Malle, B. F. (2009). *The timing of blame and intentionality: Testing the moral bias hypothesis*. Poster presented at the annual meeting of the Society of Philosophy and Psychology, Indiana University, Bloomington, IN.
- Guglielmo, S., & Malle, B. F. (2011). *Mind over morality: Mental-state inferences (still) guide moral judgments*. Manuscript submitted for publication.
- Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review*, 103, 336–355. doi:10.1037/0033-295X.103.2.336
- Hassin, R. R., Aarts, H., & Ferguson, M. J. (2005). Automatic goal inferences. *Journal of Experimental Social Psychology*, 41, 129–140. doi:10.1016/j.jesp.2004.06.008
- Holbrook, J. (2006). *The time course of social perception: Inferences of intentionality, goals, beliefs, and traits from behavior* (Unpublished doctoral dissertation). University of Oregon, Eugene, OR.
- Ickes, W., & Cheng, W. (2011). How do thoughts differ from feelings? Putting the differences into words. *Language and Cognitive Processes*, 26, 1–23. doi:10.1080/01690961003603046
- Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.
- Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219–266). New York, NY: Academic Press.
- Kalish, C. W., & Shiverick, S. M. (2004). Children’s reasoning about norms and traits as motives for behavior. *Cognitive Development*, 19, 401–416. doi:10.1016/j.cogdev.2004.05.004
- Kunda, Z., Davies, P. G., Adams, B. D., & Spencer, S. J. (2002). The dynamic time course of stereotype activation: Activation, dissipation, and resurrection. *Journal of Personality and Social Psychology*, 82, 283–299. doi:10.1037/0022-3514.82.3.283
- Lewis, P. T. (1995). A naturalistic test of two fundamental propositions: Correspondence bias and the actor-observer hypothesis. *Journal of Personality*, 63, 87–111. doi:10.1111/j.1467-6494.1995.tb00803.x
- Malle, B. F. (2004). *How the mind explains behavior: Folk explanations, meaning, and social interaction*. Cambridge, MA: MIT Press.
- Malle, B. F. (2005). Three puzzles of mindreading. In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 26–43). New York, NY: Guilford Press.
- Malle, B. F. (2008). The fundamental tools, and possibly universals, of social cognition. In R. M. Sorrentino & S. Yamaguchi (Eds.), *Handbook of motivation and cognition across cultures* (pp. 267–296). New York, NY: Elsevier/Academic Press.
- Malle, B. F. (2010). The social and moral cognition of group agents. *Journal of Law and Policy*, 20, 95–136.
- Malle, B. F., & Guglielmo, S. (2011). Are intentionality judgments fundamentally moral? In C. Mackenzie & R. Langdon (Eds.), *Emotions, imagination, and moral reasoning: Macquarie monographs in cognitive science* (pp. 275–293). Philadelphia, PA: Psychology Press.
- Malle, B. F., & Hodges, S. D. (Eds.). (2005). *Other minds: How humans bridge the divide between self and others*. New York, NY: Guilford Press.
- Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, 33, 101–121. doi:10.1006/jesp.1996.1314
- Malle, B. F., & Knobe, J. (2001). The distinction between desire and intention: A folk-conceptual analysis. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 45–67). Cambridge, MA: MIT Press.
- Malle, B. F., Knobe, J., & Nelson, S. E. (2007). Actor-observer asymmetries in explanations of behavior: New answers to an old question. *Journal of Personality and Social Psychology*, 93, 491–514. doi:10.1037/0022-3514.93.4.491
- Malle, B. F., & Pearce, G. E. (2001). Attention to behavioral events during interaction: Two actor-observer gaps and three attempts to close them. *Journal of Personality and Social Psychology*, 81, 278–294. doi:10.1037/0022-3514.81.2.278
- Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students’ daily social environments and natural conversations. *Journal of Personality and Social Psychology*, 84, 857–870. doi:10.1037/0022-3514.84.4.857
- Mitchell, J. P. (2009). Inferences about mental states. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 1309–1316. doi:10.1098/rstb.2008.0318
- Molden, D. C. (2009). Finding meaning in others’ intentions: The process of judging intentional behaviors and intentionality itself. *Psychological Inquiry*, 20, 37–43. doi:10.1080/10478400902744295
- Monroe, A. E., & Reeder, G. D. (2011). Motive-matching: Perceptions of intentionality for coerced action. *Journal of Experimental Social Psychology*, 47, 1255–1261. doi:10.1016/j.jesp.2011.05.012
- Neurobehavioral Systems. (2011). *Presentation software program* (Version 14.9). San Francisco, CA: Neurobehavioral Systems.
- O’Laughlin, M. J., & Malle, B. F. (2002). How people explain actions performed by groups and individuals. *Journal of Personality and Social Psychology*, 82, 33–48. doi:10.1037/0022-3514.82.1.33
- Povinelli, D. J., & Preuss, T. M. (1995). Theory of mind: Evolutionary

- history of a cognitive specialization. *Trends in Neurosciences*, 18, 418–424. doi:10.1016/0166-2236(95)93939-U
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1, 515–526. doi:10.1017/S0140525X00076512
- Ramsay, J. T. (2003). *The development of personality trait attribution in children: The importance of theory of mind* (Unpublished doctoral dissertation). University of Toronto, Ontario, Canada.
- Read, S. J., Jones, D. K., & Miller, L. C. (1990). Traits as goal-based categories: The importance of goals in the coherence of dispositional categories. *Journal of Personality and Social Psychology*, 58, 1048–1061. doi:10.1037/0022-3514.58.6.1048
- Read, S. J., & Miller, L. C. (2005). Explanatory coherence and goal-based knowledge structures in making dispositional inferences. In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 124–139). New York, NY: Guilford Press.
- Read, S. J., & Monroe, B. M. (2009). Must judgments about intentionality precede dispositional inference? *Psychological Inquiry*, 20, 66–72. doi:10.1080/10478400902794571
- Reeder, G. D. (2009). Mindreading: Judgments about intentionality and motives in dispositional inference. *Psychological Inquiry*, 20, 1–18. doi:10.1080/10478400802615744
- Reeder, G. D., Vonk, R., Ronk, M. J., Ham, J., & Lawrence, M. (2004). Dispositional attribution: Multiple inferences about motive-related traits. *Journal of Personality and Social Psychology*, 86, 530–544.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173–220). New York, NY: Academic Press.
- Ross, L., & Nisbett, R. E. (1991). *The person and the situation*. New York, NY: McGraw-Hill.
- Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology*, 55, 87–124. doi:10.1146/annurev.psych.55.090902.142044
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum.
- Schult, C. A. (2002). Children's understanding of the distinction between intentions and desires. *Child Development*, 73, 1727–1747. doi:10.1111/1467-8624.t01-1-00502
- Skowronski, J. J., Carlston, D. E., & Hartnett, J. (2008). Spontaneous impressions derived from observations of behavior: What a long, strange trip it's been (and it's not over yet). In N. Ambady & J. J. Skowronski (Eds.), *First impressions* (pp. 313–333). New York, NY: Guilford Press.
- Smith, E. R. (1984). Attributions and other inferences: Processing information about the self versus others. *Journal of Experimental Social Psychology*, 20, 97–115. doi:10.1016/0022-1031(84)90015-5
- Smith, E. R., & Miller, F. D. (1983). Mediation among attributional inferences and comprehension processes: Initial findings and a general method. *Journal of Personality and Social Psychology*, 44, 492–505. doi:10.1037/0022-3514.44.3.492
- Snodgrass, S. R. (1976). The development of trait inference. *Journal of Genetic Psychology*, 128, 163–172.
- Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences*, 1124, 208–224. doi:10.1196/annals.1440.012
- Tomasello, M. (1998). Social cognition and the evolution of culture. In J. Langer & M. Killen (Eds.), *Piaget, evolution, and development*, *The Jean Piaget symposium series* (pp. 221–245). Mahwah, NJ: Erlbaum.
- Tracy, J. L., & Robins, R. W. (2008). The automaticity of emotion recognition. *Emotion*, 8, 81–95. doi:10.1037/1528-3542.8.1.81
- Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, 93, 239–257. doi:10.1037/0033-295X.93.3.239
- Trope, Y. (1989). Levels of inference in dispositional judgment. *Social Cognition*, 7, 296–314. doi:10.1521/soco.1989.7.3.296
- Uleman, J. S. (2005). On the inherent ambiguity of traits and other mental concepts. In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 253–267). New York, NY: Guilford Press.
- Uleman, J. S., Saribay, S. A., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, 59, 329–360. doi:10.1146/annurev.psych.59.103006.093707
- Van der Cruyssen, L., Van Duynslaeger, M., Cortoos, A., & Van Overwalle, F. (2009). ERP time course and brain areas of spontaneous and intentional goal inferences. *Social Neuroscience*, 4, 165–184. doi:10.1080/17470910802253836
- Van Duynslaeger, M., Van Overwalle, F., & Verstraeten, E. (2007). Electrophysiological time course and brain areas of spontaneous and intentional trait inferences. *Social Cognitive and Affective Neuroscience*, 2, 174–188. doi:10.1093/scan/nsm016
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72, 655–684. doi:10.1111/1467-8624.00304
- Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, 35, 245–275. doi:10.1016/0010-0277(90)90024-E
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, 17, 592–598. doi:10.1111/j.1467-9280.2006.01750.x
- 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, 237–252. doi:10.1037/0022-3514.47.2.237
- Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69, 1–34. doi:10.1016/S0010-0277(98)00058-4
- Wright, J. C., & Mischel, W. (1987). A conditional approach to dispositional constructs: The local predictability of social behavior. *Journal of Personality and Social Psychology*, 53, 1159–1177. doi:10.1037/0022-3514.53.6.1159
- Wyer, R. S., & Carlston, D. E. (1979). *Social cognition, inference, and attribution*. London, England: Routledge.
- Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, 133, 273–293. doi:10.1037/0033-2909.133.2.273

(Appendix follows)

Appendix

Table A1
Appendix 1: Stimulus Behaviors

Untailored behaviors	Goal-tailored behaviors	Trait-tailored behaviors
u1. The woman sets the dinner table for four guests.	g1. Josh's wife frequently annoys him, and he thinks the time has come	t1. The plumber slips an extra \$50 into his wife's purse.
u2. The student shelves some journals two at a time.	g2. Kate is on her way from the bus to the supermarket.	t2. The receptionist steps in front of the old man in line.
u3. The woman sweeps the floor in the apartment hallway.	g3. Mel sits depressed and alone drinking liquor.	t3. The librarian carries the old woman's groceries across the street.
u4. The man cleans one of the dirty outside windows.	g4. The boy walks fast to the counter of the supermarket.	t4. The tailor picks his teeth during dinner at the fancy restaurant.
u5. On their dinner date, he orders an expensive wine.	g5. The father holds a spoon and tells his boy, "Even the Ninja turtles like vegetables."	t5. The farmer prepares to spray paint derogatory graffiti on the building.
u6. He asks her, "Are you doing anything Friday evening?"	g6. The girl compares tools at the hardware store.	t6. The professor has his new neighbor over for dinner.
u7. While walking, he takes out his cigarettes and a lighter.	g7. The man with the luggage goes to Denver.	t7. The butcher writes a letter to the editor about air pollution.
u8. He packs his swimsuit and sun protection and heads out the door.	g8. The student is riding his bicycle to the university as fast as he can.	t8. The accountant takes the orphan to the circus.
u9. The child asks, "What's at the end of the universe?"	g9. The toddler puts on the pajamas and turns off the light.	t9. The successful filmmaker gives his ailing mother \$20 a month.
u10. The parents and their two kids all get in the van.	g10. The woman connects the garden hose and walks towards the car.	t10. The secretary solves the mystery halfway through the book.
u11. A girl does two sit-ups with a fitness ball under her legs.	g11. While passing the pet shop the girl tells her father that everyone in her class has a puppy.	t11. The sailor leaves his wife with 20 pounds of laundry.
u12. A man pours himself a cup of coffee in the kitchen.	g12. She sits down and turns on the tub faucet.	t12. The carpenter stops his car and motions the pedestrian to cross.

Note. Of untailored behaviors, u1–u4 were verbal translations of videos developed by Baldwin, Baird, Saylor, and Clark (2001), and u5–u12 were created by the present authors. All trait-tailored behaviors were taken from Winter and Uleman (1984, p. 241). Goal-tailored behaviors g1–g11 were adapted from Hassin et al. (2005, pp. 138–139), and g12 was created by the present authors.

Received August 22, 2011
 Revision received November 14, 2011
 Accepted November 15, 2011 ■