**THIS MANUSCRIPT IS CURRENTLY UNDER REVIEW**

# First Impressions Emerge Through Dynamic Trajectories of Social Inference

Samuel A. W. Klein[1], John Andrew Chwe[2], Jonathan B. Freeman[1*]

[1]Department of Psychology, Columbia University, NY, USA

[2]Lirio, TN, USA

*Correspondence concerning this article should be addressed to Jonathan B. Freeman, Columbia University, 3022 Broadway, New York, NY, 10027, United States. Email: jon.freeman@columbia.edu

**Author Contributions**: S.A.W.K. and J.A.C designed research; S.A.W.K. and J.A.C performed research; S.A.W.K. contributed new reagents/analytic tools; S.A.W.K. analyzed data; and S.A.W.K. and J.B.F. wrote the paper.

**Competing Interest Statement**: There are no competing interests to disclose.

**Classification**: Social Sciences (major), Psychological and Cognitive Sciences (minor)
**Keywords**: first impressions; social cognition; person perception; natural language processing; spontaneous inference

Data, analysis code, and stimulus materials are available on OSF: https://osf.io/uytzb.

**Abstract**

Social impressions are often measured as stable endpoint judgments following a social encounter, despite naturally arising from temporally unfolding streams of social information (e.g., actions, speech, facial and bodily expressions). Here, we tested whether impressions emerge through progressively constructed, path-dependent trajectories of inferences made during social encounters. Study 1 (N = 404) established that impressions integrate incrementally during an encounter, rather than being assembled retrospectively at its conclusion. Impressions tracked moment-to-moment along a single dimension (e.g., warmth) closely corresponded to endpoint judgments of the same "get-to-know-you" videos truncated at matched timepoints. Studies 2–3 (N = 456) then examined how impressions are constructed from spontaneous inferences across multiple dimensions as an encounter unfolds. Participants generated free-response descriptions while viewing videos, allowing impressions to be modeled as trajectories through high-dimensional semantic space. Across an encounter, successive spontaneous inferences demonstrated incremental integration, whereby later inferences were increasingly semantically tethered to prior ones. Moreover, the unique temporal-semantic trajectories these inferences traced predicted downstream personality judgments and social decisions. Study 4 (N = 215) then offered the causal test of this trajectory framework by experimentally reordering the same video content presented to participants. Reordering reshaped impression trajectories and changed the inferences drawn from identical segments of content depending on their preceding content; order-induced differences in trajectory structure, in turn, predicted downstream differences in judgments and decisions. Our findings advance a view of impressions not as static accumulations of encoded social information, but as temporally-structured trajectories reflecting the evolving construction of social meaning.

**Significance Statement**

First impressions shape consequential decisions about whom we trust, befriend, hire, or support, yet the process by which they emerge from initial social encounters has remained elusive. Here, we tracked real-time impressions during naturalistic “get-to-know-you” videos and modeled them as trajectories through high-dimensional semantic space. Successive inferences became increasingly organized by the accumulating context of prior inferences, and the unique trajectory structure of evolving impressions predicted downstream social judgments and decisions. Experimentally reordering the video content reshaped these trajectories and the inferences drawn from identical behaviors. Together, these findings reveal that lasting first impressions are shaped by a process of real-time meaning construction during the course of an initial social encounter.

## Introduction

Forming impressions is the lifeblood of how we navigate the social world, guiding everyday decisions about who to vote for, befriend, or hire. Research has predominantly treated impressions as end-state representations measured only after they are formed. Surprisingly little work has examined how impressions actually take shape during real-time social encounters.

Impression formation has long been considered an integrative process, in which multiple pieces of social information combine into a unified representation of a person (1–3). Classic studies show that the order in which written lists of traits or behavioral descriptions are encoded shapes what impressions eventually emerge, revealing primacy effects and schema-driven assimilation (1, 4–10). Later work on impression updating similarly examined how newly presented social information is incorporated into an existing representation (11–15). Yet dominant paradigms have largely presented perceivers with discrete, researcher-selected traits or behavioral descriptions and measured impressions after encoding or probed patterns in memory (7, 8, 10). Such work has established that impressions are sensitive to prior information and the sequence in which written information is encoded, but offers limited insight into how impressions actually take shape during a social encounter.

In real-world social encounters, speech, nonverbal vocal cues, and facial and bodily expressions unfold together over time, spontaneously eliciting rich inferences about many traits, mental states, and other characteristics (16–21). Rather than simply updating a judgment along a single predefined dimension (e.g., *competence*), or a select few (e.g., *warmth*, *competence*), perceivers may move through a sequence of multidimensional construals as the encounter proceeds (e.g., *charming* → *distracted* → *incompetent*). Yet virtually nothing is known about how these spontaneously generated inferences become organized over time and coalesce into the lasting "first impressions" that persist once an initial social encounter ends.

Research on narrative comprehension offers a useful framework for addressing this problem. When comprehending an unfolding story or movie, it has long been known that perceivers do not wait until its conclusion to construct meaning. They continuously build and update a schema-based model that represents the narrative's temporal, spatial, and causal state of affairs

(22–26). New information is interpreted in relation to this evolving context, such that comprehension of a story or movie reflects a path-dependent process of meaning construction, rather than a retrospective aggregation of encoded elements.

By analogy, impression formation may be understood as an evolving trajectory of social inferences through high-dimensional semantic space. In this view, our lasting impressions may be shaped not only by the aggregate set of construals inferred during a social encounter, but also the unique temporal and semantic structure of those construals. As the encounter proceeds, earlier construals would progressively constrain the interpretation of later information, causing impression trajectories to stabilize. The resulting temporal-semantic trajectory would, in turn, help explain lasting judgments and social decisions above and beyond the overall content inferred.

Consider a perceiver who initially construes another person as quiet and reserved. As the encounter unfolds, the person's measured response to a question may be interpreted as thoughtful, ultimately leading them to seem shy but with surprising intellectual depth. Another perceiver, who initially construed the person's behavior as smooth and performative, may instead interpret the same measured response as calculated, ultimately believing the person to be impressive but inauthentic. The lasting impressions following the encounter diverge because the meaning of each behavior is interpreted within an evolving context established by prior construals.

Such path-dependent dynamics cannot be recovered from endpoint judgments alone, which collapse the sequence of evolving construals into a single summary evaluation. Recent work has begun to add much-needed temporal resolution to the impression formation process, using continuous tracking methods to demonstrate that single-trait impressions shift with unfolding verbal, visual, and vocal information during video viewing (20, 27). Yet person perception is unlikely to be confined to a select set of predefined trait dimensions. When perceivers are free to describe individuals in their own words, impressions expand substantially in breadth and dimensionality (28, 29). Capturing the unfolding structure of impressions as they are naturally experienced therefore requires methods that can track how a multidimensional person representation develops in real time.

Conceptualizing impressions as an unfolding trajectory, or social "narrative," yields three core predictions. First, impressions should exist as progressively integrated representations throughout an encounter, rather than being assembled only retrospectively once it ends. Second, successive inferences should become increasingly organized by the evolving context established by prior construals, producing greater semantic coherence as an encounter proceeds. Third, the particular structure of these trajectories should shape lasting judgments and social decisions above and beyond the aggregate content inferred. The present studies were designed to test the fundamental principles of this trajectory framework.

Study 1 directly tested the premise of incremental integration by comparing continuously tracked trait impressions with the consolidated judgments made by independent perceivers after viewing the same "get-to-know-you" videos truncated at matched timepoints. Studies 2 and 3 then extended the framework beyond single, predefined trait dimensions. While viewing videos, participants spontaneously typed descriptors of the person whenever an inference sprang to mind. These descriptors were modeled as trajectories of successive inferences through high-dimensional semantic space before testing whether they increasingly cohere to prior inferences and whether the "narrative" structure of these trajectories explains final impressions and downstream decisions. Study 4 offers a causal test to the trajectory framework by experimentally reordering the same video content, holding the overall presented information constant. If earlier construals shape how later information is interpreted, shuffling the video should alter the inferences drawn from identical video segments and the trajectories those inferences trace; in turn, such shifts in trajectory structure should shape a perceiver's lasting impressions and decisions.

## Results

### *Study 1*

In a real-time judgment-task, participants (N = 202) viewed 10 videos, for which they were randomly assigned to one of seven trait dimensions and made moment-to-moment trait assessments (0-100) using *up* and *down* arrow keys. The seven traits comprised the Big Five and Big Two dimensions in social perception: *openness, agreeableness, conscientiousness,*

*neuroticism, extraversion, warmth,* and *competence* (30, 31). In an endpoint-judgment task, an independent sample of participants (N = 202) viewed the same videos, but each truncated at one of 16 randomly assigned timepoints ranging from approximately 6–100% of full duration (Figure 1). Following each viewing, participants entered in free-response descriptors, whereby a minimum of two single-word descriptors were entered about the target, after which they provided trait ratings (11-point scale) along all seven trait dimensions. Free-response text entries were preprocessed (e.g., lowercasing, lemmatization) and represented as vectors in 300-dimensional word embedding space via the pre-trained FastText model (32) (see *Materials & Methods*).

Across Studies 1–4, applying more conservative filtering approaches (e.g., socially-relevant words only) to free-response descriptors did not meaningfully alter the findings, nor did statistical controls for mere primacy effects. Additional analyses ruled out that the pattern of results in these studies are reducible to mere shifts in valence rather than high-dimensional socially-relevant semantics (see SI Appendix).

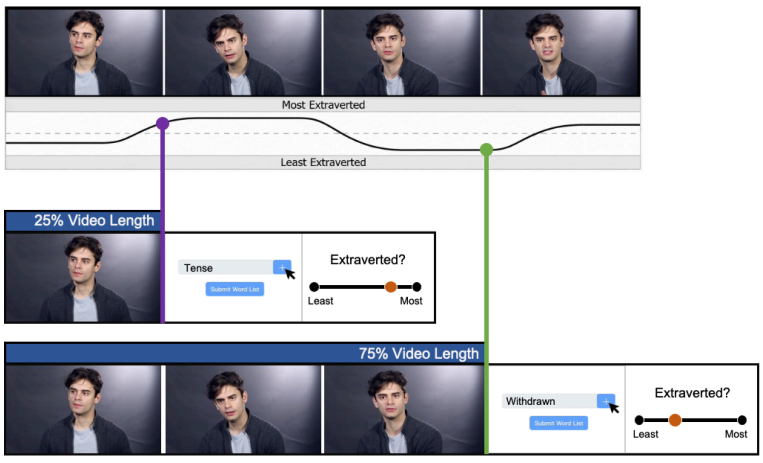


**Figure 1. Real-Time Judgment and Endpoint Judgment Tasks in Study 1.** Schematic illustration of the real-time judgment task (top) and endpoint judgment task (bottom), completed by independent participants. In the real-time judgment task, participants provided moment-to-moment ratings of each target on one assigned trait dimension, while viewing each full video. In the endpoint-judgment task, different participants viewed the same videos truncated at randomized timepoints (e.g., 25% or 75% video length) or in full before providing free-response descriptors and scalar ratings for the same seven traits.

Interrater reliability in the two tasks was moderate-to-strong, commensurate with prior work (33). Moment-to-moment trait assessments showed Spearman-Brown split-half reliabilities of .52–.93, and endpoint ratings for full-length video-viewing showed intraclass correlation coefficients of .56–.84 (SI Appendix Study 1 Supplementary Analyses; Tables S2 & S3).

We first assessed how impressions stabilized with increasing video exposure. Trait judgments made after viewing truncated videos increasingly corresponded (via Pearson $r$ on seven-dimension trait profiles) with judgments made after full-length video viewing, an effect better characterized as logarithmic (vs. linear; $\Delta$AIC = -9.6), $b = 0.08$, $SE = 0.01$, $t(1,738) = 6.65$, $p < .001$ (Figure 2b). The free-response descriptors entered after each cutoff point corroborate this pattern, also showing an increasing correspondence (via cosine similarity between average descriptor embeddings) with descriptors entered after full-length video viewing, which was again strongly logarithmic (vs. linear; $\Delta$AIC = -16.3), $b = 0.02$, $SE < 0.01$, $t(1,737) = 9.32$, $p < .001$. Thus, endpoint judgments at each video's cutoff points increasingly reflected judgments made after full-length video viewing, converging rapidly toward a relatively stable impression that continued to be refined as the encounter unfolded.

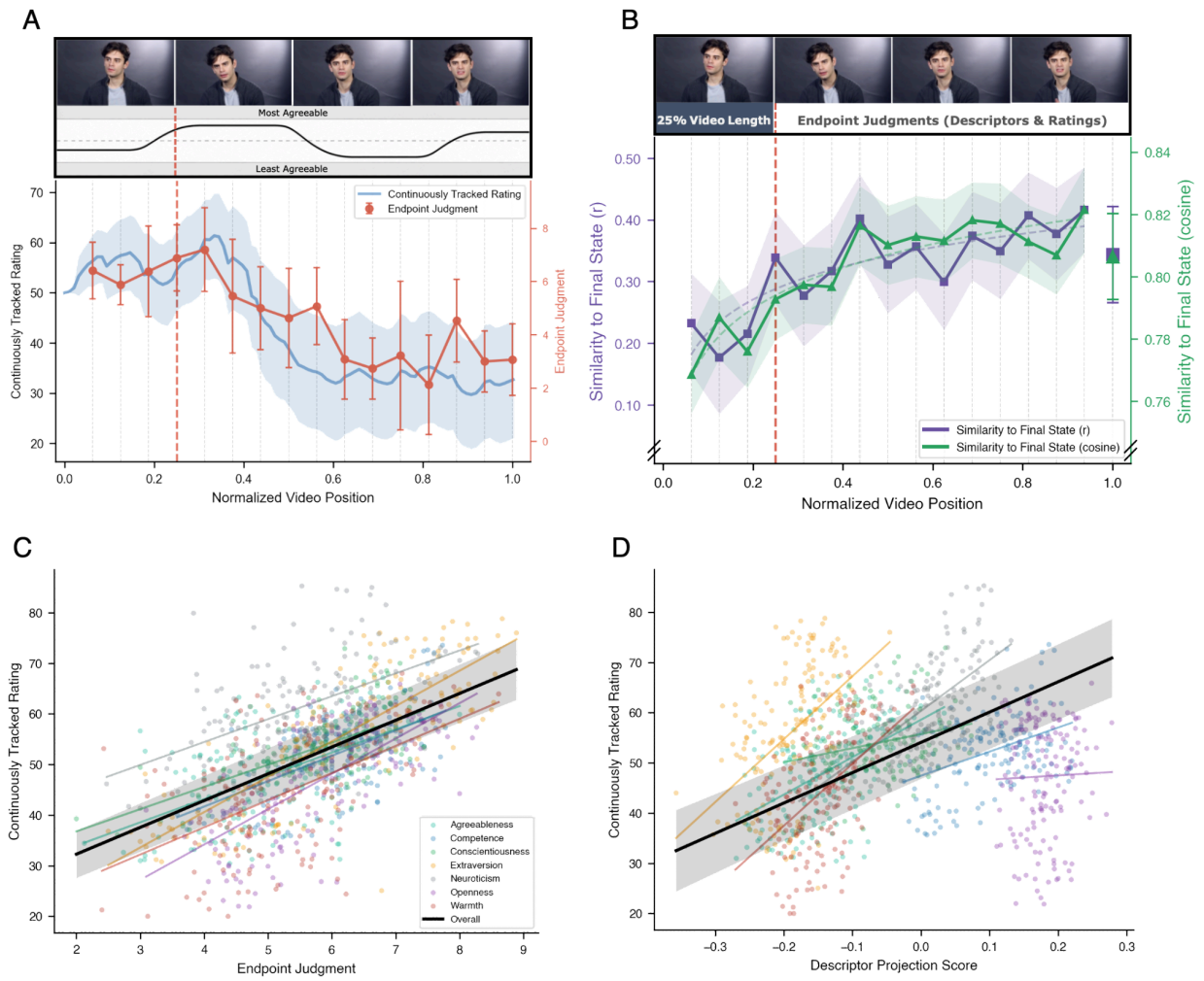


**Figure 2. Evidence for Real-Time Integration in Study 1. a)** Continuously tracked agreeableness ratings for a representative video (blue; ribbon indicates 95% CI) closely corresponded to endpoint judgments from an independent sample that viewed truncated versions of the same video (orange; means ± 95% CI), plotted along normalized video position. **b)** As viewing duration increased, endpoint trait judgments (purple, Pearson *r*) and free-response descriptors (green; cosine similarity) increasingly resembled final impressions after full-length viewing (squares; Normalized Video Position = 1.0). Ribbons indicate 95% CI. **c-d)** Continuously tracked impressions were predicted at matched timepoints by endpoint trait ratings (**c**) and free-response descriptors' projections to those same trait dimensions (**d**). Black lines show model fits with ribbons indicating 95% CIs; colored markers and lines reflect trait-specific data and illustrative OLS fits.

**Evolving Impressions Reflect Real-Time Integration.** Since endpoint judgments were made with unlimited time, they can be assumed to reflect a fully consolidated impression based on the encounter up until that point. We therefore tested whether the moment-to-moment ratings reflect

ongoing integration by examining their association to endpoint judgments, matched at the various cutoff points (e.g., momentary agreeableness ratings compared to endpoint agreeableness ratings, after viewing 25% of the video; Figure 1).

Indeed, regressing moment-to-moment trait ratings onto endpoint trait judgments at matched cutoff timepoints revealed significant correspondence, $b = 5.46$, $SE = 0.21$, $t(1,103) = 25.70$, $p < .001$ (Figure 2c). The free-response descriptors entered after each video were additionally projected onto latent dimensions of the same seven traits (see SI Appendix), yielding continuous scores indexing their trait-relevant content. Descriptor projections were significantly associated with endpoint trait ratings, $b = 7.08$, $SE = 0.16$, $t(13,776) = 43.42$, $p < .001$, indicating convergent validity. Regressing moment-to-moment trait impressions onto this more unconstrained form of endpoint judgments also showed significant correspondence, $b = 63.26$, $SE = 5.16$, $t(966) = 12.26$, $p < .001$ (Figure 2d). Nonparametric permutation testing, which shuffled endpoint judgments across cutoff points to break their sequential alignment, confirmed that these effects arose from the genuine temporal sequence observed rather than spuriously from a similar vocabulary of descriptors participants used for each video (both $p$s = .001; see SI).

***Study 2***

Study 1 provided evidence of an ongoing integration process driving real-time social impressions, but did so by constraining real-time impressions to one of seven trait dimensions. To capture how impressions unfold more naturally across multiple dimensions, Study 2 introduces a novel paradigm that tracks participants' full, high-dimensional social inference space as it evolves during an encounter. Participants (N = 150) watched the same 10 videos from Study 1, but here were asked to enter single-word, free-response descriptors whenever they noticed any new characteristic about the person, pausing the video to do so. Entries were required at the first frame and immediately after the video ended, with all intervening entries self-initiated. Multiple entries could be inputted with any pause of the video (Figure 3a). At the end of the video, following free-response entries, participants made final impressions on the seven trait dimensions used in Study 1.

Across Studies 2–4, and following standard preprocessing, participants made an average of 4–5 entries and entered 8–13 descriptors per video. Their responses predominantly described personality traits, followed by mental states, with appearance and social categories encompassing smaller proportions of responses (see SI Appendix for details).

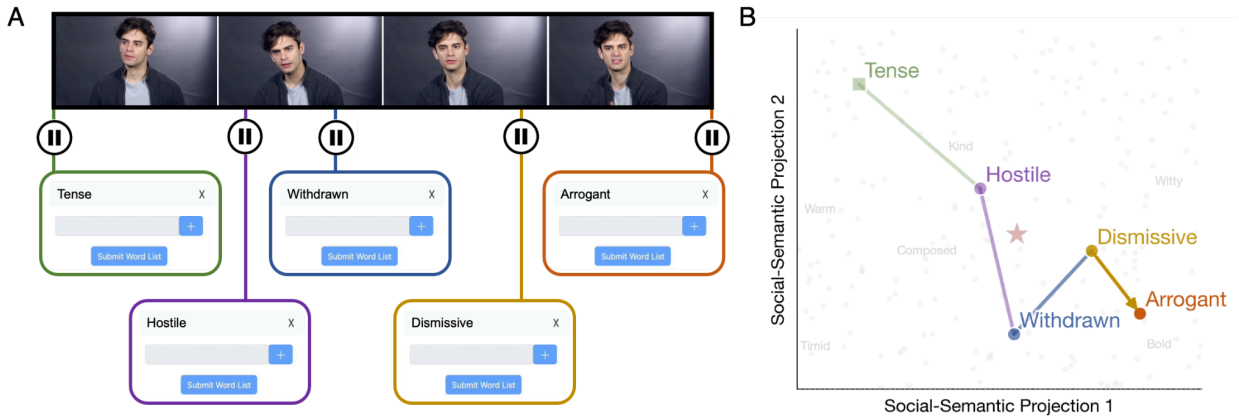


**Figure 3. Spontaneous Free-Response Descriptor Paradigm in Studies 2–4. a)** Schematic illustration of the free-response descriptor task used in Studies 2–4, in which participants pause the video at self-selected moments to enter spontaneous, single-word descriptors (e.g., *tense* → *hostile* → *withdrawn* → *dismissive* → *arrogant*). **b)** These time-stamped entries can be represented as a trajectory of social inferences evolving through a high-dimensional semantic space (via a pre-trained 300-dimension word-embeddings model), visualized here as projections in reduced two-dimension semantic space. Transparent colors indicate earlier entries; opaque colors indicate more recent entries. The star indicates average semantic content across the trajectory's five descriptors (i.e., centroid between the five points).

Each entered word was mapped onto semantic space using pretrained word embeddings (300-dimensional vectors), such that at each moment a participant's current impression state was defined by the embedding of the most recent entry. When multiple descriptors were entered at the same pause, their embedding vectors were averaged to define the impression state at that entry timepoint. The similarity between any two impression states was quantified using cosine similarity, providing a measure of how closely related two impressions are in semantic meaning.

This approach treats impression formation as a trajectory through a high-dimensional semantic space. If we consider this space as projected into two dimensions, each point reflects the perceiver's current impression state and distances between points reflect the semantic similarity

between those impression states (Figure 3b). In this manner, we can track real-time impression formation during a social encounter as a path through all possible semantic inferences over time. Study 2 tests whether participants' traversal through semantic space reflects an ongoing construction of stable social inferences that set the stage for lasting impression outcomes.

**High-Dimensional Social Inferences Integrate Over Time.** If impressions are dynamically integrated over time, each successive inference should be more tethered with the evolving trajectory defined by earlier inferences, progressively constraining the interpretation of subsequent information during the encounter. Therefore, we first examined whether each descriptor entered during the video is more semantically similar to all preceding descriptors than expected by chance.

For each spontaneous entry ($j$) after the first, indexed by its ordinal position in the sequence of $k$ total entries, $j \in \{2,\ldots,k\}$, we calculated the average pairwise cosine similarity between the descriptor(s) entered at that entry point and each preceding entry in the sequence $(1,\ldots, j-1)$, indexing the semantic coherence between each new social inference and those preceding it. This semantic coherence was compared against a null distribution of 1,000 random contexts, each constructed by drawing $j-1$ descriptors at random from the complete vocabulary of descriptors entered about that target across participants, computing the cosine similarity between each randomly drawn descriptor and the current entry, and averaging those similarities. For example, at a participant's 5$^{th}$ entry point during a video (e.g., *genuine*), we calculated the cosine similarity between that entry and each of the participant's four preceding entries (e.g., *warm* → *curious* → *thoughtful* → *sincere*), before averaging those pairwise cosine similarities and subsequently comparing them against 1,000 null draws of other participants' entries for that target (e.g., *assertive* → *shy* → lively → *focused*). This analysis yielded a semantic coherence score ($z$) at each entry $j \geq 2$, reflecting the standardized distance of the observed coherence from the null distribution (Figure 4a). Observed entries with multiple descriptors were averaged before computing their semantic coherence to prior entries.

The semantic coherence for a participant's observed impression sequence was significantly stronger than chance (random shuffles of comparable word sequences), $b = 1.22$, $SE = 0.07$, $t(31) = 18.30$, $p < .001$ (Figure 4b). To characterize how this semantic coherence evolved over time,

we modeled semantic coherence scores as a function of their step in the sequence ($j$). As in Study 1, this function was strongly logarithmic (vs. linear; ΔAIC = -20.46), with coherence growing significantly and logarithmically over time, $b = 0.11$, SE = 0.02, $t(7,192) = 4.86$, $p < .001$. That is, the largest gains in semantic coherence occurred early in the sequence, with progressively smaller increases thereafter. This result suggests that perceivers quickly settle into a region of high-dimensional semantic space, after which subsequent inferences increasingly gravitate toward this emerging representation, producing progressively stabilized yet continuously evolving impressions.

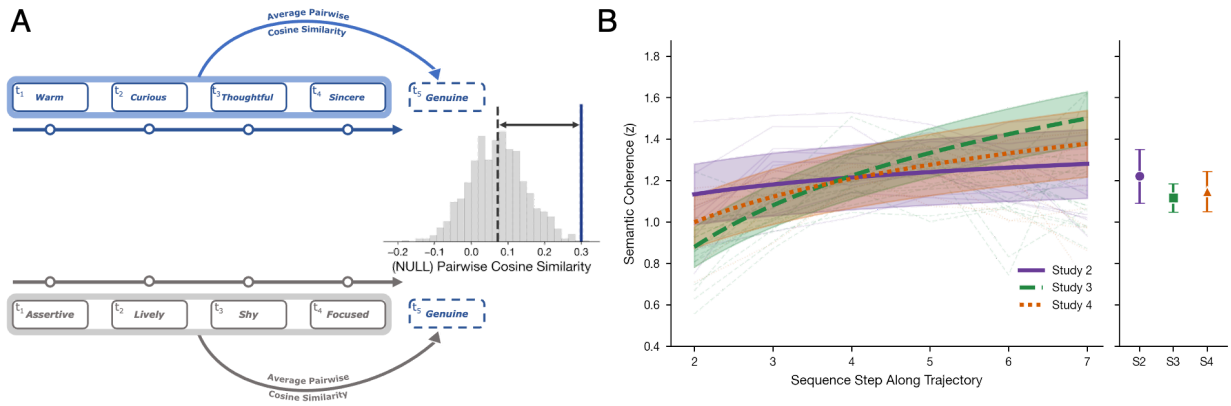


**Figure 4. Semantic Coherence Analyses in Studies 2-4. a)** Schematic of the semantic coherence computation (see Results). The observed coherence (top) was calculated as the average pairwise cosine similarity between a newly generated descriptor (e.g., $j = 5$) and each prior descriptor in the sequence (blue). Observed coherence was evaluated against a null distribution ($n = 1{,}000$; inset histogram) of mean similarities between the new descriptor and randomly drawn words from the video-specific vocabulary (bottom). **b)** Semantic coherence scores ($z$), where null hypothesis of baseline vocabulary use is $z = 0$ (dashed line), as a function of sequence step along the trajectory for Studies 2–4 (solid purple, dashed green, and dotted orange, respectively), showing a logarithmic increase, with model fits and 95% CI bands overlaid on individual video-level trajectories; intercept-only point estimates and 95% CIs testing coherence against chance for each study (right).

**Temporal and Semantic Organization of Construals Predicts Final Impressions.** Having established that successive construals become increasingly organized by prior inferences, we next examined whether the structure of these evolving trajectories is reflected in the impressions

that persist after the encounter. The trajectory framework predicts that participants whose inferences evolve in more similar ways over time should form more similar final impressions, above and beyond the aggregate, order-independent content of those inferences. Using representational similarity analysis (RSA) (34–36), we therefore tested whether similarities in participants' temporal-semantic trajectories predict corresponding similarities in their endpoint trait judgments.

For each target video, we constructed a Final Impressions dissimilarity matrix (DM) over all pairwise combinations of participants, reflecting the extent to which the seven-trait final impressions profiles were more similar or dissimilar (Pearson distance, $1 - r$).

Additionally, an Impression Trajectory DM, for each target video, captured how similarly a pair of participants' trajectories traversed high-dimensional semantic space. To that end, trajectories were first normalized to 100 time-bins and linearly interpolated between successive entries. Trajectory dissimilarity was then quantified using Dynamic Time Warping (DTW) with a cosine distance cost function, to preserve the trajectory's ordinal structure while accommodating for potential variability in pacing (e.g., one perceiver arriving at the same inference a few seconds earlier than another) (37). As a result, two trajectories were considered more similar when they followed comparable progressions through semantic space (e.g., *awkward* → *reserved* → *thoughtful*), even if the exact timing or number of descriptors differed. Finally, we constructed a Content-Only DM for each video as a control, capturing the cosine distance between each pair's aggregate semantic meaning (i.e., the mean embeddings vector across all descriptors in a trajectory). Unique dissimilarities under the DMs' diagonals were vectorized and their associations were subsequently analyzed.

Dissimilarities in the Impression Trajectory DM significantly predicted corresponding dissimilarities in the Final Impressions DM, controlling for dissimilarities in the Content-Only DM, $b = 0.13$, $SE = 0.03$, $t(60{,}490) = 4.61$, $p < .001$. That is, the extent to which the impression trajectories for any pair of participants traversed semantic space in a more similar manner over time for a target video predicted a corresponding similarity in the final impressions made by those participants, even when controlling for the two trajectories' aggregate semantic content,

(Figure 5c). The Content-Only DM also predicted the Final Impressions DM, $b = 0.63$, $SE = 0.03$, $t(60{,}170) = 18.88$, $p < .001$.

These results suggest a distinct role for the unique temporal-semantic organization of successive inferences in shaping final impressions, above and beyond the content of those inferences. To provide a more stringent test, we conducted nonparametric permutation testing. For each of 1,000 permutations, descriptor embeddings were randomly reassigned across observed timestamps within each participant-video trajectory, such that a trajectory containing *warm* at 10-s and *amused* at 22-s could become *amused* at 10-s and *warm* at 22-s. This method preserves the words produced and the time when entries were made, but destroys the unique narrative arc of the observed sequence of inferences. We then recomputed the DTW cosine distance between the resulting permuted trajectories. The effect of trajectory dissimilarity on final-impression dissimilarity significantly exceeded this permutation-based null ($p = .001$), indicating that the observed temporal order of inferences is critical for predicting downstream impressions.

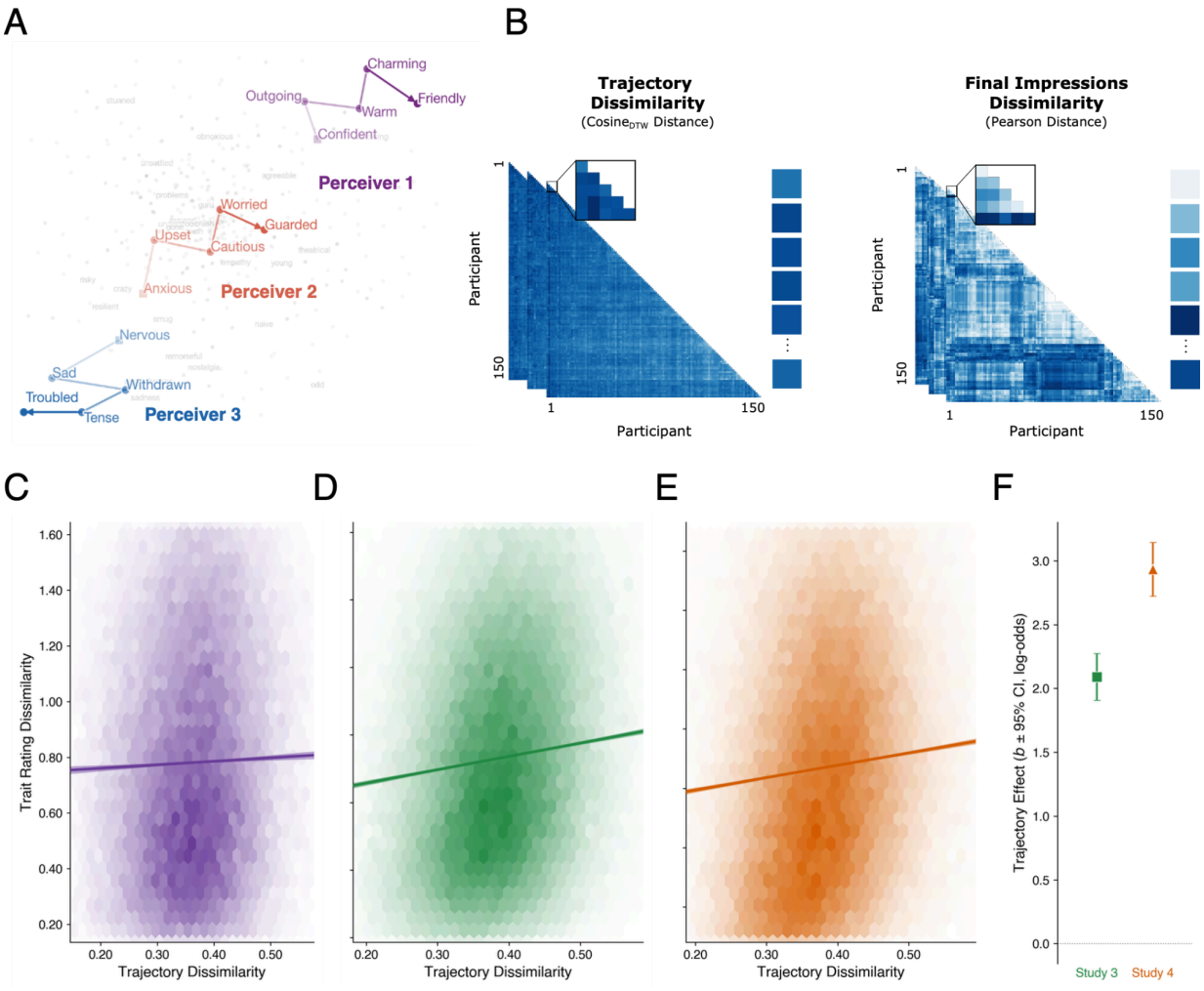


**Figure 5. Representational Similarity Analyses in Studies 2-4. a)** Schematic illustration of impression trajectories. Three perceivers' impression trajectories for the same video are visualized in reduced two-dimensional embedding space; Perceivers 1 and 2 share more similar temporal organization to each other than to Perceiver 3. Transparent colors indicate earlier entries; opaque colors indicate more recent entries. **b)** For each video, Impression Trajectory dissimilarity matrices (DMs; left) indexed pairwise trajectory dissimilarity (DTW, cosine distance cost), and Final Impressions DMs (right) indexed pairwise endpoint-rating dissimilarity (Pearson distance). **c–e)** Pairwise trajectory dissimilarity predicted final trait-rating dissimilarity in Study 2 (**c**), Study 3 (**d**), and Study 4 (**e**), controlling for Content-Only dissimilarity. Lines show mixed-effects model fits; hexbin densities show pairwise observations with the central 95% of the trait-dissimilarity range. **f)** Trajectory dissimilarity also predicted recruitment-decision disagreement in Studies 3 and 4. Markers reflect model point estimates ± 95% CI in log-odds; the dotted line indicates no effect.

***Study 3***

Study 3 (N = 306) replicated Study 2's findings with another video stimulus set and tested their implications for social decision-making. Participants served as recruiters, watching a randomly assigned sample of 10 of 18 videos (≥132 participants per video) in which MBA students delivered professional self-introductions for companies they were interested in joining (38). Participants spontaneously entered free-response descriptors throughout their viewing of each video, followed by personality trait ratings and subsequently deciding whether to recruit the candidate for an interview.

Replicating Study 2, semantic coherence for a participant's observed impression sequence was significantly stronger than chance, $b = 1.12$, $SE = 0.04$, $t(194) = 31.78$, $p < .001$ (Figure 4b), and was better characterized by a logarithmic rather than linear function ($\Delta AIC = -59.75$), with coherence growing significantly and logarithmically over time, $b = 0.50$, $SE = 0.02$, $t(9{,}865) = 20.79$, $p < .001$.

Further replicating Study 2, pairs of participant dissimilarities in each video's Impression Trajectory DM predicted corresponding dissimilarities in the Final Impressions DM, controlling for dissimilarities in the Content-Only DM, $b = 0.39$, $SE = 0.02$, $t(190{,}100) = 23.60$, $p < .001$ (Figure 5d). The Content-Only DM also predicted the Final Impressions DM, $b = 0.74$, $SE = 0.02$, $t(192{,}600) = 40.85$, $p < .001$. Permutation testing once again confirmed that this effect arose due to the particular ordering of inferences along impression trajectories ($p = .001$).

**Temporal and Semantic Organization of Construals Predicts Decisions to Recruit.** To explore the impact of these impression dynamics on social decision-making, we constructed a Recruitment Decision DM per video (0 = agreement, 1 = disagreement), using a logistic model to reveal that participant dissimilarities in trajectory structure significantly predicted decision disagreements, controlling for dissimilarities in aggregate semantic content, $b = 0.18$, $SE = 0.01$, $z = 23.84$, $p < .001$ (Figure 5f). Dissimilarities in aggregate semantic content also predicted decision disagreements, $b = 0.25$, $SE = 0.01$, $z = 33.62$, $p < .001$. Nonparametric permutation testing confirmed that the effect of trajectory dissimilarities was due to the particular ordering of inferences along impression trajectories ($p = .001$). A mediation analysis further suggested that while impression trajectories' paths through semantic space indirectly shape recruitment

decisions through their impact on final trait impressions, they also more directly shape these decisions in ways not captured by endpoint trait ratings alone (SI Appendix).

***Study 4***

Building on the correlational findings of Studies 1-3, Study 4 (N = 215) experimentally tested whether impressions unfold as path-dependent trajectories, such that the meaning assigned to any segment of social information depends on the construals that precede it. We divided each of eight videos from Study 3 into four main segments, placing boundaries at natural sentence breaks where participants in Study 3 had relatively infrequently paused to enter descriptors (thereby minimizing the risk that segmentation would split a moment likely to inspire spontaneous inferences) (Figure 6a). The introductory and concluding sentences remained fixed, while the four intervening segments were presented either in their original order or in one of nine reordered sequences (Figure 6b; see SI Appendix for greater detail on video segmentation). Thus, presentation order was manipulated while overall video content was held constant. Participants then completed the same task as in Study 3, allowing a direct test of whether order causally changes the specific inferences perceivers draw and the trajectory those inferences trace, and whether such changes carry forward to final impressions and decisions.

Semantic coherence for participants' observed impression sequences again exceeded chance, $b$ = 1.15, $SE$ = 0.05, $t(22.3) = 23.08$, $p < .001$ (Figure 4b), and grew logarithmically over the sequence, $b = 0.30$, $SE = 0.03$, $t(6{,}170.6) = 10.65$, $p < .001$ ($\Delta$AIC = −36.74 vs. linear).

Pairs of participant dissimilarities in trajectory structure again predicted corresponding final-impression dissimilarities, $b = 0.42$, $SE = 0.02$, $t(152{,}800) = 22.55$, $p < .001$ (Figure 5e), and decision disagreements, $b = 2.93$, $SE = 0.11$, $z = 27.25$, $p < .001$ (Figure 5f), controlling for dissimilarities in aggregate content. Dissimilarities in aggregate content also predicted final-impression dissimilarities, $b = 0.62$, $SE = 0.02$, $t(156{,}200) = 31.33$, $p < .001$, and decision disagreements, $b = 3.01$, $SE = 0.12$, $z = 26.20$, $p < .001$. Nonparametric permutation testing once again confirmed these effects of trajectory structure depended on the genuine inference sequence ($p$s = .001). Participants therefore generated impression trajectories that became increasingly semantically related even when the stimulus itself no longer unfolded in its natural order. This

suggests that the coherence effects observed across the studies are not merely inherited from the narrative structure of the stimulus, but are actively constructed by perceivers as they integrate each new piece of information with their evolving impression.

**Order Shuffling Reshapes Impression Trajectories.** We first tested directly whether the same social information acquired different meaning depending on the sequence that preceded it. Mean embeddings of the descriptors generated for an identical segment of a video were more similar for pairs of participants whose preceding segments had the same ordered sequence than those who had a different ordered sequence for that video, $b = 0.021$, $SE = 0.001$, $t(196{,}045) = 27.23$, $p < .001$ (Figure 6c), suggesting that earlier segments created an evolving context that changed the social meaning construed in later ones.

Converging trajectory-level analyses showed that these local differences accumulated into broader changes in the paths participants traced through semantic space. Trajectory dissimilarity (DTW, cosine distance; as in previous RSAs) was greater between participants who viewed the same video content in a different order than those who viewed the content in identical order, $b = 0.001$, $SE < 0.001$, $t(147{,}343) = 2.54$, $p = .011$, an effect that survived a nonparametric permutation test of the order labeling ($p = .016$). However, because participants within the same order encountered the same segments at the same moments, this result could partly reflect shared stimulus timing rather than differences in construal. To resolve this, we reassigned each descriptor to the position of its corresponding segment in the original unshuffled video and recomputed trajectory dissimilarities (thereby removing differences in stimulus timing while preserving the inferences generated). If each segment were interpreted independently of the context established by earlier segments, the greater similarity among participant pairs within the same viewing order than between different viewing orders should disappear after realignment. Instead, within-order participant pairs remained more similar than between-order participant pairs, $b = 0.001$, $SE < 0.001$, $t(146{,}903) = 2.18$, $p = .029$, an effect that continued to survive permutation testing ($p = .032$).

Finally, we used mediation analyses to test whether order-induced changes in trajectory structure carried forward to final impressions and decisions. Order produced no detectable overall (i.e., total) effect on final-impression dissimilarity or decision disagreement ($p$s $\geq .333$). However, the

indirect pathway was significant: participant pairs who viewed different orders showed greater trajectory dissimilarity (a = 0.01, $p$ = .006), which, in separate models, subsequently predicted greater dissimilarity in final impressions (indirect = 0.0012, 95% CI [<0.001, 0.002]; Figure 6e) and greater decision disagreement (indirect = 0.007, 95% CI [0.001, 0.012]; Figure 6f). Analyses using segment-level content as the mediator also showed the same pattern (SI Appendix).

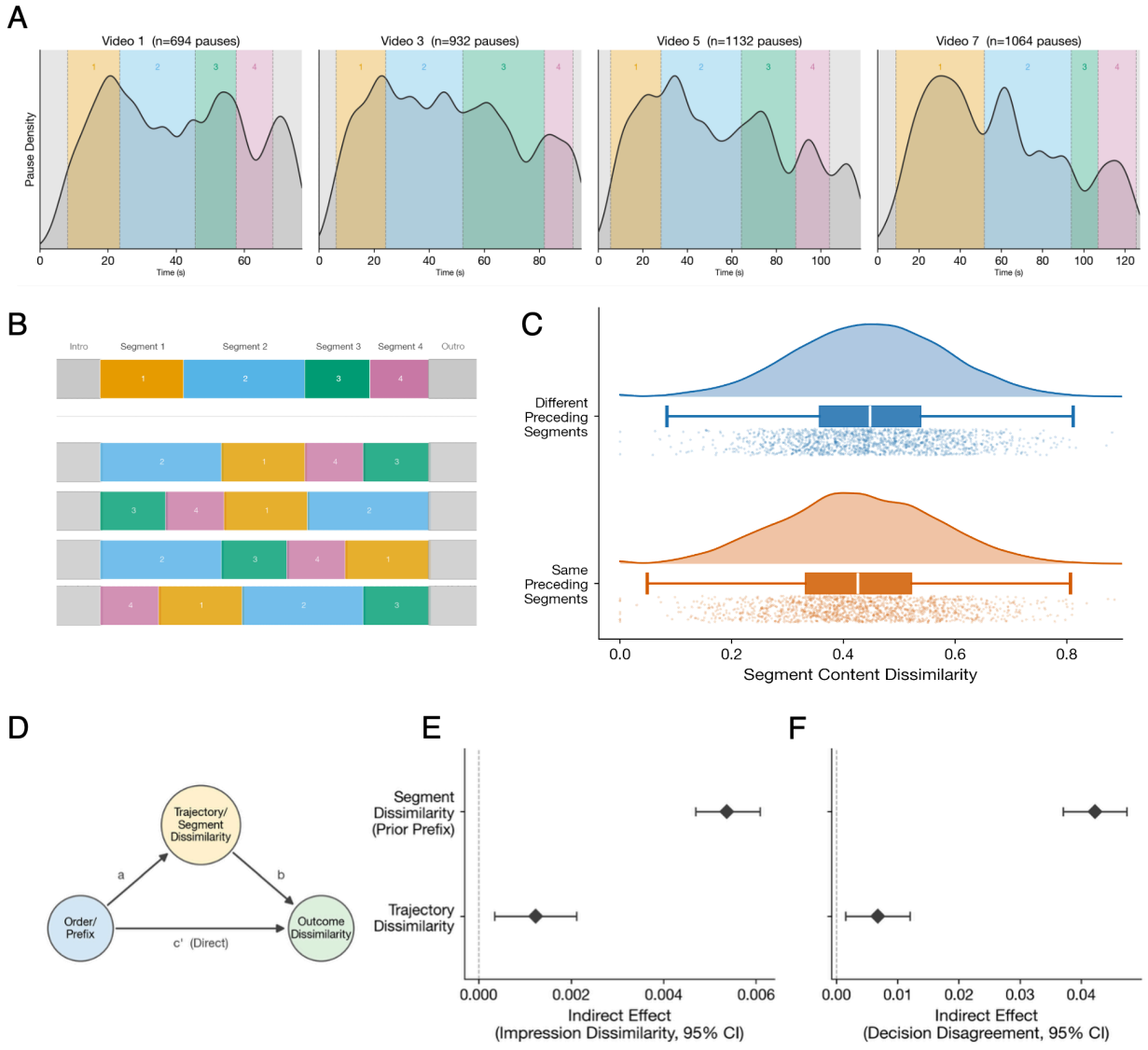


**Figure 6. Order Shuffling Reshapes Impression Trajectories in Study 4. a)** Pause-density distributions for four representative videos, showing the fixed introductory and concluding segments (gray) and four reorderable body segments (colors). Segment boundaries were placed at natural sentence breaks where relatively few Study 3 participants paused to enter descriptors. **b)** Schematic of the original segment order and four example derangements; participants viewed

either the original order or one of nine derangements. **c)** Distributions of segment content dissimilarity for pairs of participants whose preceding segments followed a different order (blue) versus the same order (orange). Boxes show the median and interquartile range; points show subsampled individual pairwise dissimilarities. Identical segments of each video produced more dissimilar descriptions when participants viewed different rather than the same preceding segments. Point estimates show model coefficients ± 95% CIs; the dashed line indicates no effect. **d)** Schematic of the Study 4 mediation models. **e–f)** Indirect effects of order or prior-prefix differences through trajectory or segment dissimilarity on final-impression dissimilarity (**e**) and recruitment-decision disagreement (**f**). Points show indirect-effect estimates ± 95% CIs; dashed lines indicate no effect.

## Discussion

Across four studies, we show that first impressions are progressively constructed as path-dependent trajectories during initial social encounters. In Study 1, real-time trait impressions closely corresponded to endpoint judgments when encounters were truncated at matched timepoints, showing that impressions exist as incrementally integrated states during an encounter. Studies 2–4 broadened this account by moving beyond single, predefined trait dimensions and using spontaneous free-response descriptions to map the high-dimensional sequence of inferences perceivers generate over time. The findings reveal that successive inferences increasingly cohere to an accumulating history of prior inferences and suggest that the particular temporal-semantic structure of these sequences shape downstream impressions and decisions. Across studies, these effects reflected higher-dimensional semantic organization, rather than mere shifts in valence or primacy effects. Study 4, in particular, tested this framework experimentally: reordering the same video content shifted the trajectories participants traced and the inferences drawn from identical social information, with these changes corresponding to differences in downstream impressions and decisions.

Classic research has shown that changing the order of experimenter-provided, written traits or behaviors can alter the impression ultimately reported (1, 39). Here, by measuring the sequential generation of perceivers' own spontaneous inferences, our findings reveal a distinct form of sequence dependence in impression formation. Rather than treating traits or behaviors as fixed inputs whose influence varies with serial position, our results suggest that each inference helps establish the interpretive context in which subsequent information acquires social meaning. In

Study 4, identical video segments elicited different spontaneous inferences depending on the sequence that preceded them, reshaping the trajectories through which impressions and decisions emerged. Early information may therefore matter for forming impressions not simply because it receives greater weight or establishes a schema that is thereafter uniformly applied (1, 5), but because it redirects the evolving path along which later social information is understood.

The logarithmic stabilization of impression trajectories, in particular, suggests that this path becomes increasingly constrained without becoming fixed: impressions rapidly settle into a region of semantic space while continuing to evolve through more gradual refinement. This pattern of continuous integration through high-dimensional semantic space is consistent with research on narrative comprehension showing that perceivers actively integrate episodic details into an evolving schema during real-time experience (22, 24, 25). Just as viewers of a movie or readers of a story continually interpret new information relative to an evolving representation of a narrative's structure, the current findings suggest that spontaneous social inferences are interpreted in the context of an accumulating history of their own prior construals. Impression formation may therefore be best understood as a particular instantiation of a domain-general process of dynamic meaning construction, in which successive inferences constrain the interpretive space. This is consistent with recent work showing that free recall of a naturalistic experience (e.g., movie) reproduces from memory the narrative's geometric structure in semantic space (40). Likewise, endpoint impressions in the present work preserve the joint temporal and semantic structure of successive social inferences, reflecting not merely a static summary but how a person "came to be understood" across the encounter, and explaining why more similar impression trajectories arrived at more similar endpoint impressions and social decisions, beyond their shared aggregate content. Together, these findings offer, to our knowledge, the first account of how spontaneous social inferences become organized over time into enduring first impressions.

Beyond helping explain how impressions form, this framework may speak to why people come away with markedly different views of the same person. Decades of research have documented substantial idiosyncrasy in how particular perceivers construe particular individuals (41–43). Yet because prior work focusing on idiosyncrasy has operated on endpoint judgments, it has left

unresolved how differences among perceivers emerge as a function of the real-time dynamics of impression formation. Here, perceivers varied in the endpoint impressions and decisions they made about the same people, and these differences were predicted by idiosyncratic variations in their evolving impression trajectories. It is possible that moment-to-moment construals may differ across perceivers from the outset, with initially small differences compounding over time. Indeed, our findings suggest that early divergence could be progressively amplified as successive inferences increasingly cohere to each perceiver's evolving context. Perceivers may also vary in how strongly they integrate each new social inference to the semantic context established by their prior inferences: whereas some perceivers may strongly constrain how they interpret later behavior on their earlier construals, others perceivers may more readily depart from the emerging schema (3, 44–48). Additionally, perceivers may differ in how they relate social concepts—for instance, linking *reserved* with *thoughtful* versus *aloof*—such that even identical inferences early on activate different expectations of future behavior, sending impressions along increasingly divergent paths (49–51). Trajectory-based approaches may expose multiple, non-mutually exclusive mechanisms that give rise to idiosyncrasies in lasting judgments and decisions.

The present findings also bolster recent high-dimensional models of social impressions, which propose that impressions unfold by converging on stable "attractor" regions within semantic space (52). Our findings extend this framework, which has largely focused on rapid impressions (e.g., split-second face impressions), to more naturalistic and temporally-extended social encounters, suggesting that ongoing impressions cohere around an attractor region of semantic space as an encounter proceeds. An open question for future work is how these dynamics unfold in longer or repeated encounters containing substantially new, unexpected, or conflicting information (11–15). Under such conditions, impressions may potentially undergo larger-scale reorganization or rapid "phase transitions" between attractor regions (53), followed by renewed stabilization as bottom-up and top-down information integrates toward a newly emerging representation (2, 18, 54–57).

Although the free-response paradigm provides a window into evolving construals, it does not offer a "pure" measure of spontaneous inference. Verbalization requires introspective access to an inference and its mapping onto a linguistic response, processes that may themselves shape

underlying representations (58, 59). Several research traditions have used similar verbalization methods (e.g., think-aloud protocol) to capture ongoing interpretations and provide a closer approximation of spontaneous cognition than retrospective or predefined-response measures can offer (59–62). Future work directly comparing response modalities could help clarify the spontaneous nature of the observed inferences in our studies. Additionally, while research on impressions has traditionally focused on inferring personality traits, growing evidence suggests that inferences of transient mental states drive inferences of enduring traits, and vice-versa (63). In line with this, here we retained all socially-relevant descriptors, but future work could clarify how trait and state inferences may differentially contribute to evolving impression trajectories.

The present research generates several novel predictions for future work. Our demonstration that post-encounter impressions preserve the structured trajectory of construals during the encounter suggests that later memory of a target may be similarly biased toward that trajectory, with recall preferentially reflecting trajectory-consistent information (40), which future work could address. Moreover, the free-response paradigm could be extended to dyadic interactions, exploring the dynamic coupling between two conversation partners' evolving trajectories, as impressions guide behavior that in turn shapes the other's inferences. Characterizing the dynamics of this coupling may provide useful insights for the burgeoning literature on conversation (64).

Overall, the present findings suggest that social impressions are dynamically constructed trajectories through semantic space that evolve through the progressive constraint of successive inferences, with their unique structure predicting downstream social judgments and decisions. Impression formation, therefore, may be best understood not merely as an accumulation of social information, but as the evolving and sequential construction of social meaning over time.

## Materials & Methods

### *Participants*

Across all studies, independent samples were recruited via Prolific. Final samples comprised N = 202 (real-time judgment task) and N = 202 (endpoint-judgment task) in Study 1, N = 150 in Study 2, N = 306 in Study 3, and N = 215 in Study 4. Study 1's endpoint-judgment task excluded participants averaging fewer than two descriptors per video (2 excluded). Studies 2–4 applied the following exclusion pipeline: descriptors were excluded if they comprised four or more words or were absent from the *FastText* model's vocabulary; trials with fewer than three remaining unique entry timepoints were excluded; and participants were excluded if fewer than three trials remained. In Study 2, one additional participant was also manually excluded for a data-collection error. Applying these criteria yielded the following total exclusions: in Study 2, 65 of 215 were excluded (30.2%); in Study 3, 72 of 378 were excluded (19.0%); in Study 4, 16 of 231 were excluded (6.9%). Including all participants did not meaningfully alter the results in any study. Full exclusion details can be found in the SI Appendix (*Data Exclusions*), as can the demographic composition of each sample (age, race, gender; Table S28). All studies were approved by Columbia University's Institutional Review Board, and all participants provided informed consent.

### *Stimuli*

Stimuli used in Studies 1 and 2 consisted of 10 first-person monologue-style videos of unfamiliar individuals sourced from YouTube, oriented directly toward the camera and selected to depict a range of emotive audition clips by amateur actors. Videos ranged in duration from 52.52 s to 131.50 s ($M$ = 98.98 s, $SD$ = 22.70 s). Study 3 stimuli consisted of 18 videos of MBA students serving as prospective job candidates delivering professional self-introductions to companies they were interested in working for, recorded in prior work (38); participants viewed a random subset of 10 of the 18 videos. Videos ranged in duration from 49.65 to 150.02 s ($M$ = 107.60 s, $SD$ = 24.95 s).

Study 4 used a subset of eight videos from Study 3 (M = 106.95 s, SD = 20.94 s). As detailed earlier, each video was segmented into six clips: a fixed introductory clip, four body segments, and a fixed concluding clip. For each participant and video, the four body segments were independently assigned one of ten possible orders (the original order or one of nine derangements) and joined with the fixed intro and outro clips into a single gapless video. Brief fades were introduced at segment boundaries by this concatenation procedure (see Procedure for explanation of these fades provided to participants).

***Procedure***

**Endpoint-Judgment Task (Variable Video Duration; Study 1).** Each participant viewed all 10 videos, with each video truncated at a randomly assigned timepoint. To generate these timepoints, each video's duration was divided into 16 normalized timepoints, ranging from approximately 6.25% of the video's length to its full duration. Participants were assigned one per video drawn randomly without replacement from the 16 possible bins, ensuring that no two videos were truncated at the same relative timepoint within a participant's study session. Presentation order of the 10 video trials was randomized across participants.

Participants received instructions describing the trial structure and that some videos might appear to end abruptly. Following each video, participants provided free-response, single-word descriptors about the target. A minimum of two descriptors was required per target. Participants then rated the target on seven trait dimensions using 11-point slider scales (0 = *least*, 10 = *most*), each initialized to the midpoint (5): *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, *Neuroticism*, *Warmth*, and *Competence*. The first five correspond to the Big Five personality factors (30); Warmth and Competence were included as additional dimensions given their centrality to social perception (28, 31). Participants subsequently provided demographic information before completing the study.

**Real-Time Judgment Task (Full Video Duration; Study 1).** Each participant viewed all 10 videos in their full duration while continuously rating the target on a single, randomly assigned trait dimension. Participants were instructed to rate how strongly the target possessed their assigned trait moment-to-moment, updating their rating as the perceived trait changed. Each trait

was accompanied by a behaviorally anchored description (e.g., Openness: "willingness to try new things and to explore new ideas"). Participants completed a demonstration trial with the continuous tracking interface before the main task, and video presentation order was randomized.

Ratings were expressed on a 0–100 scale anchored at "Least [trait]" and "Most [trait]," initialized to the midpoint (50). Participants adjusted ratings in real time using the *up* and *down* arrow keys, with an initial step size of 2 units and an acceleration multiplier of 1.1 per successive keypress, resetting on key release. Ratings were rendered as a continuously scrolling visual trace.

**Spontaneous Free-Response Descriptor Paradigm (Studies 2–4).** Participants provided free-response descriptors of their impression of the person in each video, pausing the video for each entry. Participants were required to enter at least one descriptor on the first frame before starting the video. During viewing, participants could pause at any moment by clicking the video to enter additional descriptors that came to mind. A minimum inter-pause interval of 2 s prevented accidental consecutive pauses. Descriptors were entered one word at a time into an editable list, and playback resumed automatically after the entries were submitted. Immediately following video completion, participants entered additional descriptors, representing the final entry timepoint for trajectory analyses. In Studies 3 and 4, at least two final descriptors were required to reduce data sparsity. In Study 2, participants could remove descriptors from their complete list before rating the target; this feature was removed in Studies 3 and 4 to reproduce the findings while preventing participants from revisiting prior entries, ensuring results were not unduly influenced by the descriptor list itself. In Study 4, to explain the brief fades introduced at segment boundaries by the video-stitching procedure, participants were told that videos depicted longer interviews edited to show only key excerpts (see Stimuli).

Participants then provided endpoint impressions using 11-point slider scales (0 = *least*, 10 = *most*), each initialized to the midpoint (5). In Study 2, participants rated the target on the same seven trait dimensions as Study 1; Studies 3 and 4 added four additional trait dimensions relevant to the professional context (*Leadership Ability*, *Trustworthiness*, *Confidence*, and *Ambitiousness*), and were subsequently asked whether they would invite the candidate to interview (yes/no). The procedure repeated across videos, followed by demographic questions.

### *Data Preprocessing*

All analyses were conducted in R. Linear mixed-effects models were fit using the *lme4* package, with p-values obtained using *lmerTest* and Satterthwaite degrees-of-freedom approximation. Logistic mixed-effects models were fit using lme4. The Study 3 decision-disagreement models were fit using *glmmTMB* because the equivalent *lme4* models exceeded available memory. Nonparametric permutation tests relied on empirical p-values computed by comparing observed test statistics to corresponding null distributions.

**Moment-to-Moment Ratings Data.** Raw moment-to-moment ratings were recorded as sparse event logs, with new values stored only when ratings changed using the up or down arrow. Between changes, the rating was treated as remaining at its most recently selected value. These continuous rating time series were then averaged across participants at each timepoint within each video-trait combination.

**Free-Response Descriptors.** Free-response descriptors across all studies were spell-corrected, lemmatized, stripped of stopwords, and excluded if they comprised four or more words (reflecting full sentences rather than single-word descriptors) or were absent from a pretrained FastText embeddings model (32), which represented all remaining descriptors as 300-dimensional vectors (full pipeline and per-study exclusion counts in SI Appendix). To examine correspondence with endpoint scalar ratings in Study 1, embeddings were additionally projected onto seven trait-congruence axes, defined via seed words from established psychometric lexicons (SI Appendix).

**Impression Trajectories.** Embedding vectors for each entry's free-response descriptors were assigned to their observed timestamps and linearly interpolated between successive entries across a normalized 100-bin timeline. When multiple descriptors were entered at the same pause, embedding vectors were averaged for that entry timepoint.

### *Statistical Analysis*

**Correspondence Analyses (Study 1).** Correspondence between endpoint judgments following truncated viewing with those following full-length viewing of each video was quantified using Pearson correlations and modeled with a fixed effect of cutoff time and random intercepts for participant and video. Correspondence between moment-to-moment ratings and matched-timepoint endpoint judgments was modeled with random intercepts for video and trait, for both endpoint trait ratings and descriptor trait projections. The association between descriptor trait projections and endpoint trait ratings was modeled with random intercepts for participant, video, and trait.

**Semantic Coherence Analysis (Studies 2–4).** At each entry after the first, observed semantic coherence was calculated as the mean cosine similarity between the current descriptor entry and all preceding entries and evaluated against a null distribution generated by randomly drawing an equivalent number of descriptors from the target-specific vocabulary. Semantic coherence z-scores were standardized using the pooled null SD computed across all timepoints and draws. To test whether trajectories were significantly more coherent than chance, we fit a mixed-effects intercept-only model with crossed random intercepts for participant and video. To characterize how coherence changed over time, we fit a growth model regressing z-scores on sequence step, with the same random effects structure, comparing linear and logarithmic specifications via AIC.

**Representational Similarity Analysis (RSA; Studies 2–4).** For each video, dissimilarity matrices (DMs) indexed all pairwise combinations of participants. Impression Trajectory DMs captured the dissimilarity between participants' temporally ordered descriptor trajectories, which was computed using Dynamic Time Warping with a cosine-distance cost function and a 20% Sakoe-Chiba band. Final Impressions DMs indexed Pearson distance ($1 - r$) between participants' seven-dimensional trait-rating profiles in Study 2 and eleven-dimensional profiles in Studies 3 and 4. Content-Only DMs indexed cosine distance between the mean embedding vectors across all descriptors in each participant's trajectory, thereby capturing aggregate semantic content independent of temporal order. In Studies 3 and 4, Recruitment Decision DMs indexed pairwise disagreement in recruitment decisions (0 = same decision, 1 = different decisions). Final-impression dissimilarity and recruitment-decision disagreement were modeled as functions of trajectory dissimilarity while controlling for content-only dissimilarity. All

models included random intercepts for video and participant pair (i.e., dyad). For permutation tests, descriptor embeddings were randomly reassigned across their observed timestamps within each trajectory, preserving the descriptors produced and the timing of entries while disrupting their observed temporal order. Impression Trajectory DMs were recomputed and the corresponding mixed-effects model was refit for each of 1,000 permutations. Empirical p-values were computed as the proportion of permuted absolute t- or z-statistics exceeding the observed value, two-tailed.

**Order Shuffling Analyses (Study 4).** For each identical video segment, pairwise content dissimilarity was calculated as the cosine distance between participants' mean descriptor embeddings. Mixed-effects models tested whether segment content was more dissimilar when participants had viewed different versus identical preceding sequences, with random intercepts for video and dyad. Additional models, with the same random effects structure, were used to test whether entire impression trajectories were more dissimilar for participants assigned different rather than identical presentation orders. To remove differences attributable to stimulus timing, descriptors were additionally realigned to the corresponding segments' positions in the original unshuffled video before trajectory dissimilarities were recomputed. Permutation tests randomly reassigned order-condition labels within video to evaluate the observed same- versus different-order effect.

**Mediation Analysis (Studies 3–4).** In Study 3, we tested whether Final Impression DMs mediated the relationship between Impression Trajectory DM (X) and Recruitment Decision DM (Y), reflecting the task sequence: trajectories accumulated during video-viewing, followed by trait ratings, before decisions were finally made. In Study 4, we tested whether trajectory dissimilarity or segment-level content dissimilarity mediated the relationship between the order manipulation (X: same vs. different order; same vs. different sequence of preceding segments) and downstream dissimilarity in final impressions or decisions (Y).

Continuous outcomes (ratings dissimilarities) were modeled via linear mixed-effects models; binary outcomes (decision disagreement) were modeled via logistic mixed-effects models. All models included crossed random intercepts for video and dyad. The indirect effect was estimated via quasi-Bayesian simulation (5,000 iterations), sampling fixed effects from each model's

approximate posterior. For Study 3, counterfactual probabilities were computed with covariates held at their means and averaged across treatment conditions for binary outcomes to account for logit nonlinearity; for Study 4, indirect effects were computed as the product of independently sampled path coefficients (full implementation details in SI Appendix). The indirect effect represents the estimated change in the outcome for a one standard deviation increase in trajectory dissimilarity (Study 3) or between levels of the categorical order predictor (Study 4).

**Acknowledgements**

This work was supported in part by National Science Foundation research grant BCS-2235130 (J.B.F.). We thank Yara Studer for instrumental coding contributions.

**References**


1. S. E. Asch, Forming impressions of personality. *J. Abnorm. Psychol.* **41**, 258–290 (1946).

2. Z. Kunda, P. Thagard, Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychol. Rev.* **103**, 284–308 (1996).

3. S. J. Read, L. C. Miller, *Connectionist models of social reasoning and social behavior*, S. John, L. C. Miller, Eds. (Psychology Press, 2014).

4. N. H. Anderson, Integration theory and attitude change. *Psychol. Rev.* **78**, 171–206 (1971).

5. N. H. Anderson, Foundations of information integration theory. (1981).

6. R. M. Hogarth, H. J. Einhorn, Order effects in belief updating: The belief-adjustment model. *Cogn. Psychol.* **24**, 1–55 (1992).

7. R. Hastie, B. Park, The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychol. Rev.* **93**, 258–268 (1986).

8. M. Lichtenstein, T. K. Srull, Processing objectives as a determinant of the relationship between recall and judgment. *J. Exp. Soc. Psychol.* **23**, 93–118 (1987).

9. A. R. McConnell, S. J. Sherman, D. L. Hamilton, On-line and memory-based aspects of individual and group target judgments. *J. Pers. Soc. Psychol.* **67**, 173–185 (1994).

10. T. K. Srull, R. S. Wyer, Person memory and judgment. *Psychol. Rev.* **96**, 58–83 (1989).

11. J. Cone, M. J. Ferguson, He did what? The role of diagnosticity in revising implicit evaluations. *J. Pers. Soc. Psychol.* **108**, 37–57 (2015).

12. T. C. Mann, M. J. Ferguson, Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *J. Pers. Soc. Psychol.* **108**, 823–849 (2015).

13. P. Mende-Siedlecki, S. G. Baron, A. Todorov, Diagnostic value underlies asymmetric updating of impressions in the morality and ability domains. *J. Neurosci.* **33**, 19406–19415 (2013).

14. S. A. W. Klein, R. J. Hutchings, A. R. Todd, Revising mental representations of faces based on new diagnostic information. *Cognition* **217**, 104916 (2021).

15. M. Masi, S. Mattavelli, F. Fasoli, M. Brambilla, Cross-modal impression updating: Dynamic impression updating from face to voice and the other way around. *Br. J. Soc. Psychol.* **61**, 808–825 (2022).

16. Y. Hu, C. J. Parde, M. Q. Hill, N. Mahmood, A. J. O'Toole, First impressions of personality traits from body shapes. *Psychol. Sci.* **29**, 1969–1983 (2018).

17. C. A. Sutherland, *et al.*, Social inferences from faces: Ambient images generate a

three-dimensional model. *Cognition* **127**, 105–118 (2013).

18. J. B. Freeman, N. Ambady, A dynamic interactive theory of person construal. *Psychol. Rev.* **118**, 247–279 (2011).

19. P. McAleer, A. Todorov, P. Belin, How do you say “hello”? Personality impressions from brief novel voices. *PLoS One* **9**, e90779 (2014).

20. C. Lin, M. Thornton, Bottom-up and top-down information determinants of naturalistic trait impression updating. (2024).

21. N. Ambady, F. J. Bernieri, J. A. Richeson, “Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream” in *Advances in Experimental Social Psychology Volume 32*, Advances in Experimental Social Psychology., (Elsevier, 2000), pp. 201–271.

22. A. C. Graesser, M. Singer, T. Trabasso, Constructing inferences during narrative text comprehension. *Psychol. Rev.* **101**, 371–395 (1994).

23. J. P. Magliano, K. Dijkstra, R. A. Zwaan, Generating predictive inferences while viewing a movie. *Discourse Process.* **22**, 199–224 (1996).

24. R. A. Zwaan, G. A. Radvansky, Situation models in language comprehension and memory. *Psychol. Bull.* **123**, 162–185 (1998).

25. J. M. Zacks, N. K. Speer, K. M. Swallow, T. S. Braver, J. R. Reynolds, Event perception: a mind-brain perspective. *Psychol. Bull.* **133**, 273–293 (2007).

26. G. A. Radvansky, J. M. Zacks, Event boundaries in memory and cognition. *Curr. Opin. Behav. Sci.* **17**, 133–140 (2017).

27. J. Zaki, N. Bolger, K. Ochsner, It takes two: the interpersonal nature of empathic accuracy: The interpersonal nature of empathic accuracy. *Psychol. Sci.* **19**, 399–404 (2008).

28. G. Nicolas, X. Bai, S. T. Fiske, A spontaneous stereotype content model: Taxonomy, properties, and prediction. *J. Pers. Soc. Psychol.* **123**, 1243–1263 (2022).

29. J. Lu, C. Lin, Network models reveal high-dimensional social inferences in naturalistic settings beyond latent construct models. *Commun. Psychol.* **3**, 98 (2025).

30. L. R. Goldberg, The development of markers for the Big-Five factor structure. *Psychol. Assess.* **4**, 26–42 (1992).

31. S. T. Fiske, A. J. C. Cuddy, P. Glick, J. Xu, A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Soc. Cogn.* (2018).

32. P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, Enriching word vectors with subword

information. *Trans. Assoc. Comput. Linguist.* **5**, 135–146 (2017).

33. S. Vazire, M. R. Mehl, Knowing me, knowing you: the accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *J. Pers. Soc. Psychol.* **95**, 1202–1216 (2008).

34. N. Kriegeskorte, M. Mur, P. Bandettini, Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).

35. R. M. Stolier, E. Hehman, J. B. Freeman, A dynamic structure of social trait space. *Trends Cogn. Sci.* **22**, 197–200 (2018).

36. S. Y. Xie, R. Zheng, C. Lin, E. Hehman, A tutorial on representational similarity analysis for research in social cognition. *Soc. Cogn.* **43**, 167–193 (2025).

37. D. Berndt, J. Clifford, Using dynamic time warping to find patterns in time series. 359–370 (1994).

38. J. Schroeder, N. Epley, The sound of intellect: Speech reveals a thoughtful mind, increasing a job candidate's appeal: Speech reveals a thoughtful mind, increasing a job candidate's appeal. *Psychol. Sci.* **26**, 877–891 (2015).

39. D. L. Hamilton, L. B. Katz, V. O. Leirer, "Organizational processes in impression formation" (1980).

40. A. C. Heusser, P. C. Fitzpatrick, J. R. Manning, Geometric models reveal behavioural and neural signatures of transforming experiences into memories. *Nat. Hum. Behav.* **5**, 905–919 (2021).

41. D. A. Kenny, L. Albright, Accuracy in interpersonal perception: a social relations analysis. *Psychol. Bull.* **102**, 390–402 (1987).

42. D. A. Kenny, PERSON: a general model of interpersonal perception. *Pers. Soc. Psychol. Rev.* **8**, 265–280 (2004).

43. E. Hehman, R. M. Stolier, J. B. Freeman, J. K. Flake, S. Y. Xie, Toward a comprehensive model of face impressions: What we know, what we do not, and paths forward. *Soc. Personal. Psychol. Compass* **13**, e12431 (2019).

44. A. P. Dijksterhuis, V. K. A. D., A. W. Kruglanski, C. Schaper, Motivated social cognition: Need for closure effects on memory and judgment. *Journal of Experimental Social Psychology* **32**, 254–270 (1996).

45. D. M. Driscoll, D. L. Hamilton, R. M. Sorrentino, Uncertainty orientation and recall of person-descriptive information. *Pers. Soc. Psychol. Bull.* **17**, 494–500 (1991).

46. A. R. McConnell, Implicit theories: Consequences for social judgments of individuals. *J. Exp. Soc. Psychol.* **37**, 215–227 (2001).

47. C. Sava-Segal, C. Richards, M. Leung, E. S. Finn, Individual differences in neural event segmentation of continuous experiences. *Cereb. Cortex* **33**, 8164–8178 (2023).

48. J. J. Skowronski, Honesty and intelligence judgments of individuals and groups: The effects of entity-related behavior diagnosticity and implicit theories. *Soc. Cogn.* **20**, 136–169 (2002).

49. J. A. Brooks, J. B. Freeman, Conceptual knowledge predicts the representational structure of facial emotion perception. *Nat. Hum. Behav.* **2**, 581–591 (2018).

50. R. M. Stolier, E. Hehman, M. D. Keller, M. Walker, J. B. Freeman, The conceptual structure of face impressions. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 9210–9215 (2018).

51. R. M. Stolier, E. Hehman, J. B. Freeman, Trait knowledge forms a common structure across social cognition. *Nat. Hum. Behav.* **4**, 361–371 (2020).

52. J. B. Freeman, C. Lin, A high-dimensional model of social impressions. *Trends Cogn. Sci.* **29**, 790–801 (2025).

53. M. J. Spivey, S. E. Anderson, R. Dale, The phase transition in human cognition. *New Math. Nat. Comput.* **05**, 197–220 (2009).

54. S. J. Read, L. C. Miller, Rapist or "regular guy": Explanatory coherence in the construction of mental models of others. *Pers. Soc. Psychol. Bull.* **19**, 526–540 (1993).

55. J. B. Freeman, R. M. Stolier, J. A. Brooks, Dynamic interactive theory as a domain-general account of social perception. *Adv. Exp. Soc. Psychol.* **61**, 237–287 (2020).

56. R. J. Hutchings, E. Freiburger, M. Sim, K. Hugenberg, Racial prejudice affects representations of facial trustworthiness. *Psychol. Sci.* **35**, 263–276 (2024).

57. K. Hugenberg, S. G. Young, M. J. Bernstein, D. F. Sacco, The categorization-individuation model: an integrative account of the other-race recognition deficit. **117**, 1168 (2010).

58. J. W. Schooler, T. Y. Engstler-Schooler, Verbal overshadowing of visual memories: Some things are better left unsaid. *Cogn. Psychol.* (1990).

59. T. D. Wilson, J. W. Schooler, Thinking too much: Introspection can reduce the quality of preferences and decisions. *J. Pers. Soc. Psychol.* **60**, 181–192 (1991).

60. A. G. Greenwald, "Cognitive learning, cognitive response to persuasion, and attitude change" in *Psychological Foundations of Attitudes*, (Elsevier, 1968), pp. 147–170.

61. J. T. Cacioppo, R. E. Petty, Social psychological procedures for cognitive response assessment: The thought-listing technique. *Cognitive assessment* (1981).

62. K. A. Ericsson, H. A. Simon, How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind Cult. Act.* **5**,

178–186 (1998).

63. C. Lin, M. Thornton, Evidence for bidirectional causation between trait and mental state inferences. *J. Exp. Soc. Psychol.* **108**, 104495 (2023).

64. T. Wheatley, M. A. Thornton, A. Stolk, L. J. Chang, The emerging science of interacting minds. *Perspect. Psychol. Sci.* **19**, 355–373 (2024)