

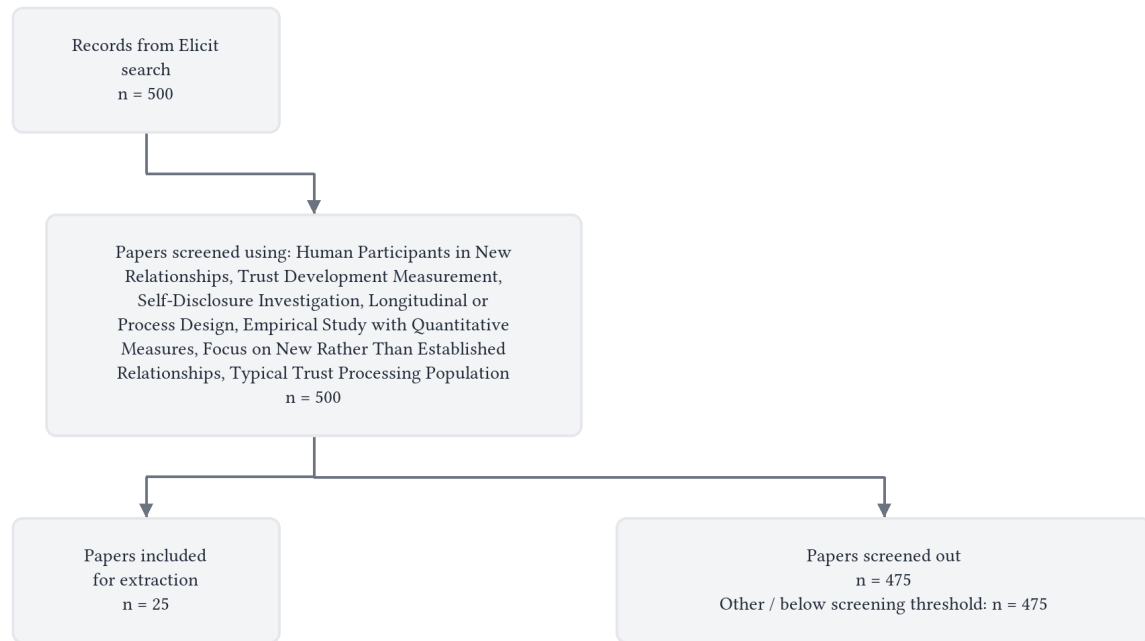
How do humans develop trust incrementally in new relationships, and what role does reciprocal self-disclosure play in deepening or withdrawing trust?

Trust develops incrementally through positive, nonlinear trajectories that progress from surface organizational cues to deeper personal disclosure, with turn-taking reciprocity in self-disclosure deepening trust via interpersonal liking and social support while self-concealment creates reciprocal cycles that withdraw trust.

Abstract

Trust develops incrementally through positive, nonlinear trajectories in most new relationships, with rapid initial growth followed by stabilization [1–3], though patterns vary substantially by relationship context. Workplace relationships show consistent nonlinear development over the first three months [1], while romantic relationships exhibit more variable patterns, with some maintaining high trust from inception rather than developing it incrementally [4]. Turn-taking reciprocity in self-disclosure—where partners alternately share at matched intimacy levels—produces higher trust than extended or non-reciprocal disclosure patterns [5, 5], operating through mediators including interpersonal liking [5], social support [6, 6], and perceived similarity [7]. However, successful disclosure unfolds through role differentiation where one partner leads intimacy escalation rather than through strict mutual reciprocity [8]. Self-concealment creates reciprocal cycles where perceived partner concealment leads to loss of trust, which then increases further concealment [9], demonstrating how non-disclosure withdraws trust. The disclosure-trust relationship is moderated by timing (early disclosure aids trust while excessive continuation harms it) [10], cultural context (positive effects appear for Americans but not Japanese or Koreans) [11], and social identity (which shapes initial trust levels but not subsequent updating patterns) [12]. Trust progression follows staged development from reliance on organizational cues to personal cues requiring deeper disclosure [1, 3], with early disclosure establishing impressions that persist even after dozens of subsequent interactions [13].

Flow Diagram



Paper search

We performed a semantic search across over 138 million academic papers from the Elicit search engine, which includes all of Semantic Scholar and OpenAlex.

We ran this query: "How do humans develop trust incrementally in new relationships, and what role does reciprocal self-disclosure play in deepening or withdrawing trust?"

The search returned 500 total results from Elicit.

We retrieved 500 papers most relevant to the query for screening.

Screening

We screened in sources based on their abstracts that met these criteria:

- **Human Participants in New Relationships:** Does this study involve human participants who are forming or developing new interpersonal relationships of any type (romantic, friendship, professional, therapeutic, etc.)?
- **Trust Development Measurement:** Does this study measure trust as a primary or secondary outcome, including trust formation, deepening, maintenance, or withdrawal?
- **Self-Disclosure Investigation:** Does this study investigate self-disclosure as a variable, whether as an intervention, exposure, or measured behavior?

- **Longitudinal or Process Design:** Does this study examine trust development over time or across multiple interactions (rather than only single-timepoint assessments)?
- **Empirical Study with Quantitative Measures:** Is this an empirical study (experimental, quasi-experimental, longitudinal observational, cross-sectional with retrospective process measures, systematic review, or meta-analysis) that provides quantitative measures of trust and/or self-disclosure?
- **Focus on New Rather Than Established Relationships:** Does this study examine relationship development phases beyond just well-established, long-term relationships (i.e., does it include initial development phases rather than focusing exclusively on established relationships of several years)?
- **Typical Trust Processing Population:** Does this study involve participants without diagnosed conditions that fundamentally alter trust processing (i.e., excludes populations with paranoid personality disorder, autism spectrum disorders where trust processing is significantly impaired, or similar trust-related disorders)?

We considered all screening questions together and made a holistic judgement about whether to screen in each paper. At abstract screening, the number of papers excluded for each primary reason was:

- **Other / below screening threshold:** n = 475

Data extraction

We asked a large language model to extract each data column below from each paper. We gave the model the extraction instructions shown below for each column.

- **Relationship Context:**

Extract details about the type and context of new relationships studied for trust development, including:

- Relationship type (romantic, workplace, friendship, online, etc.)
- How 'new' is defined (first meeting, within first weeks/months, etc.)
- Setting/context (laboratory, field, online, workplace, etc.)
- Participant characteristics relevant to new relationship formation (age, prior relationship experience, etc.)

- **Trust Development Pattern:**

Extract the specific pattern of how trust develops incrementally in these new relationships, including:

- Timeline studied (days, weeks, months)
- Shape of trust development (linear, nonlinear, stages, phases)
- Direction of change (increasing, decreasing, fluctuating)
- Speed of development (gradual, rapid initial growth, etc.)
- Critical time points or turning points in trust development

- **Self-Disclosure Behaviors:**

Extract all self-disclosure behaviors and patterns studied in relation to trust development, including:

- Types of disclosure (personal information, feelings, experiences, secrets)
- Reciprocal patterns (turn-taking, matching depth, extended disclosure, non-reciprocal)
- Timing of disclosure in relationship development
- Depth/intimacy level of disclosures
- Frequency or amount of disclosure

- **Disclosure-Trust Effects:**

Extract findings about how reciprocal self-disclosure affects trust development in new relationships, including:

- Positive effects on trust (deepening, strengthening)
- Negative effects on trust (withdrawal, deterioration)
- Conditions that determine positive vs. negative effects
- Magnitude of effects (effect sizes, strength of associations)
- Immediate vs. delayed effects of disclosure on trust

- **Trust Mechanisms:**

Extract explanations for HOW and WHY reciprocal self-disclosure influences trust development, including:

- Proposed mechanisms (vulnerability, perceived similarity, liking, social presence, etc.)
- Mediating factors between disclosure and trust
- Moderating variables that strengthen/weaken the disclosure-trust relationship
- Theoretical explanations provided by authors

- **Trust Measurement:**

Extract how trust development was measured and studied over time, including:

- Trust measures used (scales, behavioral indicators, intentions)
- Study design for capturing incremental development (longitudinal, diary, experimental)
- Number and timing of measurement points
- Whether both partners/parties were measured
- Reliability and validity of trust measurements

Results

Characteristics of Included Studies

The review included 25 studies examining trust development in new relationships across diverse contexts. Table 1 presents the key characteristics of these studies.

Study	Full text retrieved?	Relationship Context	Timeline Studied	Study Design	Primary Focus
Qiong Chen et al., 2023	No	Online interactions between strangers [5]	Two interaction phases [5]	Experimental, online [5]	Reciprocal self-disclosure patterns and trust [5]
L. V. D. Werff et al., 2018	No	Workplace, new coworkers during socialization [2]	Not specified, four waves [2]	Longitudinal field study [2]	Trust cues and disclosure intentions [2]

Study	Full text retrieved?	Relationship Context	Timeline Studied	Study Design	Primary Focus
Lisa van der Werff et al., 2017	Yes	Workplace, first three months of employment [1]	Three months [1]	Longitudinal field study [1]	Trust development during newcomer socialization [1]
L. V. D. Werff et al., 2014	No	Workplace, new coworker relationships [3]	Four waves, timing not specified [3]	Longitudinal field study [3]	Trust cues during socialization [3]
J. Ng et al., 2011	No	Student work teams [14]	Ten weeks [14]	Longitudinal observation [14]	Conflict and trust dynamics [14]
Ahmet Uysal et al., 2012	No	Romantic relationships [9]	8-10 weeks (Study 1), 14 days (Study 2) [9]	Longitudinal survey and daily diary [9]	Self-concealment and trust cycle [9]
C. A. Vanlear et al., 1987	Yes	Same-sex dyads, no prior history [15]	Six weeks [15]	Longitudinal, controlled setting [15]	Social penetration and self-disclosure [15]
M. Serva et al., 2005	No	Interacting teams in workplace context [16]	Six weeks [16]	Controlled field study [16]	Reciprocal trust between teams [16]
Young-ok Yum et al., 2005	No	Online relationships [11]	Not mentioned [11]	Cross-cultural comparison [11]	Self-disclosure and relationship development [11]
S. Altschuller et al., 2008	No	Newly formed virtual teams [17]	Not specified [17]	Laboratory experiment [17]	Trust determinants in virtual teams [17]
S. Murray et al., 2009	Yes	Romantic, newlyweds and dating relationships [18]	14 days within first months of marriage [18]	Daily diary and experiments [18]	Trust-insurance and dependence [18]
Fanzheng Yang et al., 2025	Yes	Trust between strangers [12]	Multiple stages, not timed in days/weeks [12]	Laboratory trust game [12]	Trust updating and social identity [12]
F. Krueger et al., 2007	Yes	Online interaction between strangers [19]	Short-term, multiple rounds [19]	Laboratory fMRI study [19]	Neural correlates of trust [19]

Study	Full text retrieved?	Relationship Context	Timeline Studied	Study Design	Primary Focus
Jusheng Liu et al., 2022	Yes	Online health community, patient-physician [6]	One month [6]	Cross-sectional field study [6]	Self-disclosure and physician trust [6]
H. A. Mcallister et al., 1980	No	College males, possibly romantic/friendship [20]	Not mentioned [20]	Laboratory experiment [20]	Self-disclosure intimacy and trust [20]
Pareezad Zarolia et al., 2017	Yes	Simulated financial transactions (Trust Game) [13]	Repeated interactions, no specific timeline [13]	Laboratory and online studies [13]	Indirect information and trust [13]
S. Murray et al., 2013	No	Romantic, newlyweds [21]	Three years [21]	Longitudinal study [21]	Self-protection and marital satisfaction [21]
John D. Davis et al., 1976	No	Same-sex acquaintance-ships/friendships [8]	Not explicitly mentioned [8]	Acquaintance exercise, likely laboratory [8]	Intimacy negotiation in disclosure [8]
Kari. A. Terzino et al., 2009	No	New college roommate relationships [22]	One semester [22]	Field study [22]	Self-disclosure, forgiveness, and commitment [22]
Matthew K. Miller et al., 2021	Yes	Initial interactions via video chat [23]	Single 15-minute session [23]	Laboratory, adapted to remote [23]	Media sharing vs. icebreakers for trust [23]
Qiyang Du et al., 2025	No	Workplace, newcomers and supervisors [10]	Not mentioned, multi-wave [10]	Longitudinal field study [10]	Social drinking and supervisory trust [10]
D. Bell et al., 2018	No	New sexual relationships [4]	Three years [4]	Longitudinal study [4]	Risk, trust, and protection trajectories [4]
T. Schneeberger et al., 2023	Yes	Human-SIA (Socially Interactive Agent) [24]	Single 40-50 minute session [24]	Laboratory study [24]	Closeness development with agents [24]
R. Larzelere et al., 1980	No	Romantic, intimate partners [25]	Not mentioned [25]	Not specified [25]	Dyadic trust scale development [25]

Corrie B. Whitmore et al., 2014	Yes	Roommate relationships [7]	First five weeks of semester [7]	Longitudinal field study [7]	Social exchange and trust development [7]
---------------------------------	-----	----------------------------	----------------------------------	------------------------------	---

The studies represented substantial heterogeneity in relationship contexts, timelines, and methodological approaches. Thirteen studies had full text available for detailed review. Relationship contexts included workplace relationships (five studies) [1–3, 10, 16], romantic relationships (four studies) [9, 18, 21, 25], online interactions (four studies) [5, 6, 11, 19], roommate relationships (two studies) [7, 22], and other contexts including student teams [14], virtual teams [17], sexual relationships [4], video chat introductions [23], social penetration experiments [8, 15], trust games [13], and human-agent interaction [24].

Study timelines varied considerably. Short-term studies captured trust development within single sessions [23, 24] or days [9, 18], while medium-term studies spanned weeks to months [1, 5–7, 14–16], and long-term studies followed relationships for years [4, 21]. Study designs included longitudinal field studies (ten studies) [1–4, 7, 9, 10, 21, 22], laboratory experiments (seven studies) [5, 12, 13, 17, 19, 20, 24], and other designs including daily diaries [9, 18], cross-sectional studies [6], and cross-cultural comparisons [11].

Patterns of Trust Development Over Time

Studies examining temporal patterns revealed consistent evidence of nonlinear trust development in early relationship stages. In workplace contexts, both reliance and disclosure intentions developed in positive, nonlinear patterns [1–3], with faster rates of growth during the initial month followed by stabilization between subsequent measurement periods [1]. This pattern appeared across multiple workplace studies, though specific timelines varied [2, 3].

Trust development showed increasing trajectories across diverse relationship types. In online interactions between strangers, both interpersonal trust and liking increased from the first to second interaction phase [5]. Student work teams observed time as a key driver in the transformation of trust dynamics over ten weeks [14]. Daily diary studies of romantic relationships revealed gradual trust development processes [18], though one roommate study found trust actually decreased significantly from week one to week five [7], highlighting context-specific variation.

The shape of trust development varied by measurement approach and context. Social penetration studies documented a convex quadratic trend for private-personal disclosures over time [15], suggesting an inverted U-shaped pattern where intimate disclosure peaked mid-relationship before declining. Studies using trust games found cyclical, nonlinear patterns where risk-taking actions predicted trustworthiness perceptions, which in turn predicted subsequent trust and further risk-taking behaviors in a continuous cycle over six weeks [16]. In laboratory settings with social identity manipulations, trust updating occurred asymmetrically based on initial trust levels, with low-trust individuals more responsive to positive feedback and high-trust individuals more responsive to negative feedback [12].

Critical time points emerged in several studies. The initial stages of relationship formation appeared particularly important, with early trust intentions being related to individual trust propensity [2, 3]. In newlywed relationships studied over 14 days, feelings of inferiority triggered immediate dependence-promoting behaviors aimed at maintaining trust [18]. For human-agent interactions, interpersonal closeness increased at three measurement points (after questions 12, 24, and 36), with specific trust moderating this increase [24].

Speed of development varied considerably. Some studies documented rapid initial growth in trust [1, 19], while others described gradual processes [6, 15]. Long-term studies revealed that initial self-protection levels significantly

predicted satisfaction changes over three years [21], and relationship trajectories over three years showed that most long-term relationships maintained high trust from the start rather than developing it incrementally [4].

Self-Disclosure Behaviors and Patterns

Studies examined multiple dimensions of self-disclosure relevant to trust development. Type-based categorizations included public, semiprivate, and private-personal disclosure [15], with private-personal disclosures showing the strongest association with relationship development. In online health contexts, disclosure consisted primarily of personal information measured by message length [6], while verbal and nonverbal (media-based) disclosures both contributed to trust formation in video chat settings [23].

Reciprocal patterns of disclosure emerged as critical for trust development. Turn-taking reciprocity, where partners alternately disclosed at matched intimacy levels, produced higher interpersonal trust than extended reciprocity or non-reciprocity [5]. Reciprocity at the same level of intimacy served as an interactional norm, particularly at relationship beginnings and endings [15]. However, matching was achieved not through mutual reciprocity but through role differentiation, where one partner assumed responsibility for prescribing intimacy levels while the other largely reciprocated [8].

The depth and timing of disclosure varied systematically across relationship development. Intimacy increased linearly as encounters progressed [8, 8], with partners matching their rates of increase in intimacy as well as their average intimacy levels during exercises [8]. In structured disclosure paradigms using 36 questions, intimacy increased both within and across three sets of questions [24]. Early trust was based on organizational and role-related cues, while later trust relied more on personal cues requiring deeper disclosure [1, 3].

Frequency and amount of disclosure showed important associations with trust. In workplace contexts, disclosure intentions increased over time in positive, nonlinear patterns [2, 3]. Online health communities demonstrated that increased self-disclosure by patients promoted interaction and improved satisfaction [6], with the average length of textual messages serving as a disclosure measure [6]. However, excessive disclosure over time could prove detrimental, as seen in supervisory relationships where increasing social drinking (a form of social disclosure) led to decreased trust [10].

Non-disclosure patterns also affected trust development. Self-concealment—keeping secrets from partners—created reciprocal cycles where perceived partner concealment led to loss of trust, which then led to increased self-concealment [9]. This pattern appeared in both long-term romantic relationships studied over 8-10 weeks and daily interactions over 14 days [9], suggesting the cycle operated at multiple timescales.

Effects of Reciprocal Self-Disclosure on Trust

Table 2 presents the effects of self-disclosure on trust development across studies reporting quantitative or qualitative evidence of these relationships.

Study	Disclosure Pattern	Effect on Trust	Conditions/Moderators	Strength of Effect
Qiong Chen et al., 2023	Turn-taking reciprocity vs. extended/non-reciprocity [5]	Higher trust with turn-taking [5]	Immediacy in synchronous communication [5]	Turn-taking showed higher trust than extended and non-reciprocity conditions [5]

Study	Disclosure Pattern	Effect on Trust	Conditions/Moderators	Strength of Effect
Lisa van der Werff et al., 2017	Disclosure intentions during socialization [1]	Positive effect on trust [1]	Benevolence consistently related to disclosure [1]	Not quantified [1]
L. V. D. Werff et al., 2014	Disclosure intentions over time [3]	Positive, increasing over time [3]	Shift from organizational to personal cues [3]	Not quantified [3]
J. Ng et al., 2011	Disclosures in conflict contexts [14]	Crucial for building trust [14]	Knowledge exchange, problem solving, positive communication [14]	Not quantified [14]
Ahmet Uysal et al., 2012	Self-concealment (non-disclosure) [9]	Loss of trust, reciprocal cycle of concealment [9]	Low trust led to more concealment [9]	Trust mediated concealment-concealment association [9]
C. A. Vanlear et al., 1987	Reciprocity at same intimacy level [15]	Deepened trust [15]	Social bonds from interdependent activities [15]	Reciprocity especially important at beginning and end [15]
Young-ok Yum et al., 2005	Self-disclosure breadth and depth [11]	Positive association with trust for Americans only [11]	Cultural differences [11]	Not quantified [11]
S. Altschuller et al., 2008	Self-disclosure in virtual teams [17]	Significant correlation with trust [17]	Impression formation, social presence [17]	Not quantified [17]
Jusheng Liu et al., 2022	Patient self-disclosure [6]	Positive influence on physician trust [6]	Mediated by social support [6]	Complete mediation by social support [6]
H. A. Mcallister et al., 1980	High vs. low intimacy disclosure [20]	Used to attribute trust [20]	Intimacy level as signal [20]	General measures unsuccessful [20]
Kari. A. Terzino et al., 2009	Self-disclosure in roommate relationships [22]	Strong predictor of commitment [22]	Relational self-construal and power [22]	Moderated by self-construal and power [22]
Matthew K. Miller et al., 2021	Media-based and conversational disclosure [23]	Similar trust levels across conditions [23]	Agreeableness did not moderate media sharing [23]	Media sharing increased warranting value [23]
Qiyang Du et al., 2025	Social drinking as disclosure behavior [10]	Positive initially, negative over time [10]	Timing and intensity of behavior [10]	Initial positive association, later negative trajectory [10]
T. Schneeberger et al., 2023	Structured self-disclosure (Fast Friends) [24]	Increased interpersonal closeness [24]	Moderated by specific trust [24]	Significant moderator effect of specific trust [24]

Study	Disclosure Pattern	Effect on Trust	Conditions/Moderators	Strength of Effect
R. Larzelere et al., 1980	Intimacy of self-disclosure [25]	Associated with trust, especially in longer relationships [25]	Duration of relationship [25]	Partners reciprocated trust more than disclosure depth [25]

Positive effects of self-disclosure on trust emerged consistently across contexts. Reciprocal self-disclosure was associated with increased interpersonal trust, with effects appearing in both initial and subsequent interactions [5]. In workplace settings, disclosure intentions showed positive, increasing trajectories [3], and personal trust cues like benevolence remained related to disclosure across all measurement points during socialization [1]. Online health interactions demonstrated that patient self-disclosure positively influenced trust in physicians [6], with significant coefficients indicating the strength of these associations [6].

Negative effects appeared primarily through non-disclosure behaviors. Perceived partner concealment was associated with loss of trust, and low trust in partners led to increased self-concealment from them [9]. This created a reciprocal cycle where concealment, loss of trust, and further concealment reinforced each other, slowly deteriorating relationship well-being [9]. Similarly, excessive social drinking behaviors over time—which can be viewed as a form of social disclosure—were associated with decreases in supervisory trust [10].

The conditions determining positive versus negative effects involved several factors. Turn-taking reciprocity produced higher trust than extended reciprocity or non-reciprocity, with immediacy in synchronous conversations enhancing this effect [5, 5]. In online health contexts, physicians' provision of social support completely mediated the relationship between patient self-disclosure and trust [6, 6]. For newcomer-supervisor relationships, timing proved critical: high levels of social drinking at the initial socialization stage associated positively with trust, but increases in drinking over time associated with trust decreases [10].

Cultural context moderated disclosure-trust relationships. Self-disclosure showed positive associations with trust for Americans but not for Japanese or Koreans in online relationship development [11]. Individual differences also mattered: relational self-construal and power dynamics moderated how self-disclosure predicted commitment (closely related to trust) in roommate relationships [22]. In video chat interactions, agreeableness moderated the effectiveness of icebreaker questions but not media sharing for building trust, suggesting media sharing provided a more equitable path to trust formation across personality types [23].

Immediate versus delayed effects showed mixed patterns. Some studies emphasized rapid trust formation within single sessions [23, 24], while others documented gradual processes over weeks or months [6]. The Fast Friends paradigm produced measurable increases in closeness within 40-50 minutes [24], suggesting disclosure can have immediate effects. However, longitudinal studies revealed that trust developed through both immediate and cumulative processes, with initial trust levels and subsequent experiences both shaping trajectories [12, 21].

Mechanisms Explaining Disclosure-Trust Relationships

Multiple theoretical mechanisms explained how reciprocal self-disclosure influenced trust development. Interpersonal liking emerged as a mediating factor, with reciprocal disclosure contributing to trust through the mechanism of positive interpersonal liking [5]. Social support served as another complete mediator in patient-physician relationships, where self-disclosure increased trust by promoting physicians' provision of informational and emotional support [6].

Perceived similarity functioned as both a mechanism and moderator. In roommate relationships, perceived similarity measured early in the relationship predicted trust development, with social exchange mediating this relationship [7]. The cognitive process whereby individuals perceived others as similar to themselves facilitated initial trust, which then developed through social interactions [7]. Role differentiation in disclosure represented another mechanism, where partners negotiated intimacy levels through a process in which one partner led disclosure and the other reciprocated, rather than through strictly mutual reciprocity [8, 8].

Social presence and impression formation contributed to trust in virtual contexts. In newly formed virtual teams, impression formation, public self-awareness, perceived social presence, and self-disclosure all showed significant correlations with trust [17]. These factors were recommended as design features for emergency response communication systems to promote trust among teammates [17]. Similarly, in video chat interactions, media sharing increased the warranting value of self-disclosures—making each disclosure more verifiable and thus more valuable for forming trust [23].

Neural mechanisms underlying trust development revealed distinct systems for different trust types. The paracingulate cortex was critically involved in building trust by inferring intentions and predicting behavior [19]. Conditional trust selectively activated the ventral tegmental area (linked to reward evaluation), while unconditional trust activated the septal area (linked to social attachment) [19]. This suggested that different neural systems implemented distinct trust strategies in partnerships [19].

Uncertainty reduction theory explained how social behaviors reduced uncertainty and built trust in supervisor-newcomer relationships [10]. High levels of social drinking at the initial socialization stage were positively associated with trust development, presumably by reducing uncertainty about the relationship [10]. However, excessive drinking over time weakened this relationship, suggesting a moderating variable where the benefits of uncertainty reduction had limits [10].

Knowledge exchange, problem solving, and positive social-emotional communication served as crucial mechanisms for building trust in the face of conflict [14]. Different types of conflict linked with different trust bases: task conflict with knowledge-based trust, relationship conflict with identification-based trust, and process conflict with calculus-based trust [14]. This suggested that disclosure and communication behaviors operated through distinct pathways depending on the relationship context [14].

Social identity shaped initial trust levels but did not alter updating patterns [12]. In-group favoritism increased initial trust levels, shifting more participants into high-trust categories and leading to stronger reactions to negative feedback [12]. Conversely, outgroup conditions resulted in greater initial distrust and heightened responsiveness to positive feedback [12]. This demonstrated that social identity influenced where trust started but not how it evolved through repeated interactions [12].

The trust-insurance system represented a complex mechanism in romantic relationships. When feeling inferior to partners, individuals activated implicit contingencies linking inferiority to exchange anxieties, which motivated behaviors to increase partners' dependence [18]. This system operated automatically for low self-esteem individuals but only under conscious deliberation for high self-esteem individuals [18]. The mechanism aimed to minimize power imbalances by making one partner indispensable to the other, thus ensuring trust-insurance [18].

Measurement Approaches for Trust Development

Studies employed diverse methods to capture trust development over time, as summarized in Table 3.

Study	Trust Measure Type	Study Design	Measurement Points	Both Parties Measured	Reliability/Validity
Qiong Chen et al., 2023	Self-reported and behavioral [5]	Experimental, two phases [5]	Not specified [5]	Not specified [5]	Not mentioned [5]
L. V. D. Werff et al., 2018	Intentions to rely and disclose [2]	Four-wave longitudinal field [2]	Four waves [2]	Not specified [2]	Not mentioned [2]
Lisa van der Werff et al., 2017	Behavioral Trust Inventory [1]	Four-wave longitudinal field [1]	Four time points [1]	Focused on newcomer intentions [1]	Established scales, common method bias addressed [1]
L. V. D. Werff et al., 2014	Intentions to rely and disclose [3]	Four-wave longitudinal field [3]	Four waves [3]	Not specified [3]	Not mentioned [3]
J. Ng et al., 2011	Behavioral indicators via IPA [14]	Longitudinal observation [14]	Not specified [14]	Not specified [14]	Not mentioned [14]
Ahmet Uysal et al., 2012	Not specified [9]	Two-point survey (8-10 weeks); Daily diary (14 days) [9]	2 points; 14 days [9]	Yes, couples measured [9]	Not mentioned [9]
C. A. Vanlear et al., 1987	Social Network Instrument [15]	Longitudinal [15]	Multiple during stay [15]	Likely both parties [15]	$\alpha = .85$, substantial factor loadings [15]
M. Serva et al., 2005	Risk-taking actions, trustworthiness perceptions [16]	Six-week controlled field [16]	Not specified [16]	Yes, both roles measured [16]	Not mentioned [16]
S. Murray et al., 2009	Anxiety scales, commitment questioning, dependence-promoting behaviors [18]	Daily diary [18]	14 daily entries [18]	Yes, both partners [18]	$\alpha = .79$ for inferiority anxiety [18]
Fanzheng Yang et al., 2025	Binary trust game choices [12]	Laboratory experiment [12]	Five stages: initial + 4 feedback rounds [12]	Both trustors and trustees [12]	Incentive-compatible methods [12]
F. Krueger et al., 2007	Percentage choosing to trust [19]	Multi-round trust game with fMRI [19]	Three phases: pre/during/post scanning [19]	Yes, dyads [19]	$\alpha < 0.05$ significance level [19]

Study	Trust Measure Type	Study Design	Measurement Points	Both Parties Measured	Reliability/Validity
Jusheng Liu et al., 2022	Patient increment over 1 month [6]	Cross-sectional [6]	Two time points (November-December 2020) [6]	Focused on patient choice [6]	Practical indicator [6]
Pareezad Zarolia et al., 2017	Trial-by-trial sharing, memory estimates, impression scales [13]	Within-subjects, repeated measures [13]	Five blocks plus pre/post impressions [13]	Multiple partners [13]	Consistent across studies [13]
Matthew K. Miller et al., 2021	Interpersonal Trust Scale (11 items) [23]	Between-subjects comparison [23]	One measurement post-call [23]	Yes, both partners [23]	$\alpha = .891$ [23]
Qiyang Du et al., 2025	Not specified [10]	Multi-wave longitudinal field [10]	Not specified [10]	Not specified [10]	Not mentioned [10]
T. Schneeberger et al., 2023	Specific trust (3 items), general trust scale [24]	Structured interaction [24]	Three points (after 12, 24, 36 questions) [24]	Only participants measured [24]	High Cronbach's alpha [24]
R. Larzelere et al., 1980	Dyadic Trust Scale [25]	Not specified [25]	Not specified [25]	Not specified [25]	Reliable, unidimensional, free from bias [25]
Corrie B. Whitmore et al., 2014	Measure of Trust in a Specific Peer [7]	Longitudinal weekly measures [7]	Five weekly points [7]	One party only [7]	Established scales, some overlap with social exchange [7]

Measurement approaches divided into three primary categories: behavioral indicators, self-reported scales, and intentions. Behavioral measures included risk-taking actions in trust games [12, 16, 19], trial-by-trial sharing behavior [13], patient choice of physicians [6], and interaction patterns analyzed through Bales' Interaction Process Analysis [14]. Self-reported scales included the Interpersonal Trust Scale [23], Dyadic Trust Scale [25], Behavioral Trust Inventory [1], and measures of trust in specific peers [7]. Intention-based measures captured willingness to rely on and disclose to others [1–3].

Study designs for capturing incremental development showed considerable variation. Longitudinal field studies typically employed four to five measurement waves over periods ranging from weeks to months [1–3, 7, 10]. Daily diary methods captured day-to-day fluctuations in trust over shorter periods [9, 18]. Experimental designs used single sessions with multiple measurement points [24], repeated interactions in trust games [12, 13, 19], or two-phase experimental manipulations [5]. Cross-sectional designs measured trust at two time points to infer development [6].

Timing and frequency of measurements varied substantially. Short-term studies measured trust at single time points post-interaction [23] or across three points during a structured session [24]. Medium-term studies employed weekly

measurements over five weeks [7] or multiple waves over several months [1–3]. Daily measurement approaches captured 14 consecutive days of trust-related behaviors and feelings [9, 18]. Long-term studies followed relationships for multiple years but with less frequent measurement points [4, 21].

Measurement of both partners varied across studies. Dyadic studies of romantic couples measured both partners' concealment, trust, and behaviors [9, 18]. Trust game studies necessarily measured both trustors and trustees [12, 19]. Team studies captured behaviors in both development and management roles [16]. However, many workplace socialization studies focused only on newcomers' intentions rather than measuring both newcomers and coworkers [1, 7].

Reliability and validity of trust measures showed mixed reporting. Studies using established scales reported good reliability, such as $\alpha = .891$ for the Interpersonal Trust Scale [23], $\alpha = .79$ for inferiority anxiety scales [18], and $\alpha = .85$ for social network trust measures [15]. The Dyadic Trust Scale was described as unidimensional, reliable, and relatively free from response biases [25]. Behavioral Trust Inventory studies addressed common method bias [1]. However, many studies did not report specific reliability or validity statistics [2, 3, 5, 9, 10, 14, 16].

Synthesis

The literature revealed substantial heterogeneity in how trust develops across different relationship contexts and the role of self-disclosure in this process. Rather than indicating inconsistent findings, this heterogeneity reflected systematic differences based on relationship type, disclosure patterns, measurement approaches, and temporal dynamics.

Context-Specific Trust Trajectories

Trust development patterns differed markedly by relationship context. In workplace settings during newcomer socialization, trust followed positive, nonlinear trajectories with rapid initial growth in the first month followed by stabilization [1]. This pattern held across multiple independent workplace studies [2, 3], suggesting it represented a reliable characteristic of trust formation in organizational contexts where roles, rules, and hierarchies structured interactions [3].

In contrast, romantic relationships showed more variable trajectories. Newlywed couples experienced gradual trust development intertwined with dependence-promoting behaviors [18], while a three-year study of sexual relationships found that most long-term partnerships maintained high trust from inception rather than developing it incrementally [4]. The apparent contradiction resolved when considering that these studies examined different relationship stages: the former focused on very early marriage (two to six months), while the latter identified relationship trajectories over years. The subset showing declining protection over time represented only 15% of relationships [4], with the majority maintaining stable high-trust, low-protection patterns from the start.

Roommate relationships presented yet another pattern, with trust actually decreasing from week one to week five in one study [7]. This decline occurred despite social exchange predicting trust [7], suggesting that the rewards versus costs balance in randomly assigned roommate relationships often proved disappointing as initial optimism met reality. This contrasted with workplace contexts where formal roles and professional norms may have sustained trust development even when personal chemistry was limited.

Disclosure Patterns and Trust Outcomes

The type and pattern of reciprocal disclosure proved more important than mere disclosure frequency. Turn-taking reciprocity—where partners alternately disclosed at matched intimacy levels—produced higher trust than extended

reciprocity (one partner disclosing multiple times before the other responds) or non-reciprocity [5, 5]. This effect persisted across both initial and subsequent interactions [5], with immediacy in synchronous conversations enhancing the benefit [5].

However, strict mutual reciprocity was not actually how successful disclosure unfolded in practice. Instead, partners engaged in role differentiation where one partner assumed responsibility for prescribing intimacy levels and the other largely reciprocated [8]. The more disclosing partner generally led this process [8], creating a dynamic asymmetry even while producing matched intimacy levels overall. This explained why some studies found reciprocity crucial [5, 15] while others emphasized role allocation [8]—both captured the same underlying phenomenon from different angles.

The depth-timing relationship also showed systematic patterns. Early workplace trust relied on organizational cues (roles, rules, group identification), while mature trust depended on personal cues requiring deeper disclosure [1, 3]. This staged progression from surface to depth appeared across multiple contexts. The social penetration process showed convex quadratic trends for private-personal disclosures [15], peaking mid-relationship before declining, while intimacy overall increased linearly [8]. This suggested an optimal disclosure trajectory where intimate sharing intensified during relationship consolidation but moderated once established.

Mediating Mechanisms

The pathway from disclosure to trust operated through multiple mediators depending on context. In online health communities, social support completely mediated the disclosure-trust relationship [6, 6], meaning self-disclosure only built trust insofar as it prompted physicians to provide informational and emotional support. Without that supportive response, disclosure alone did not enhance trust. This highlighted the fundamentally interactive nature of trust development through disclosure.

Interpersonal liking mediated disclosure-trust links in online stranger interactions [5], while perceived similarity mediated this relationship in roommate contexts, with social exchange serving as a secondary mediator [7]. These different mediators reflected the distinct bases of trust across relationships types. Health relationships centered on instrumental support, stranger interactions on affective responses, and roommate relationships on similarity-based expectations and ongoing exchange patterns.

Impression formation mediated the effects of indirect behavioral information on trust [13], demonstrating that trust could develop not only from direct self-disclosure but also from hearing about how partners treated others. These impressions proved remarkably persistent, influencing trust even after dozens of direct interactions that should have overridden initial impressions [13]. This suggested a primacy effect where early information—whether from disclosure or observation—established trust baselines that proved difficult to revise.

Moderating Factors

Several variables systematically moderated disclosure-trust relationships. Social identity shaped initial trust levels through in-group favoritism but did not alter how trust updated over time [12]. Those starting with high trust (due to in-group identity) reacted more strongly to negative feedback, while those starting with low trust (outgroup conditions) showed heightened responsiveness to positive feedback [12]. This asymmetry in updating meant that trust fragility versus resilience depended critically on starting conditions rather than the updating process itself.

Cultural context moderated the disclosure-trust link, with positive associations appearing for Americans but not Japanese or Koreans in online relationships [11]. This likely reflected cultural differences in communication norms and the meaning of disclosure, with direct self-revelation carrying different implications across individualistic versus collectivistic societies. The absence of an effect did not mean disclosure was unimportant in these cultures, but rather that it operated through different mechanisms or at different intimacy levels than measured.

Individual differences also moderated effects. Relational self-construal and power dynamics moderated how self-disclosure predicted commitment in roommate relationships [22], while agreeableness moderated the effectiveness of icebreaker questions but not media sharing for building trust [23]. This suggested that media sharing provided a more equitable pathway to trust formation, less dependent on personality traits that might advantage or disadvantage certain individuals in traditional conversation-based disclosure.

Timing emerged as a crucial moderator. High levels of social drinking initially associated positively with supervisory trust during newcomer socialization but increasing drinking over time associated with trust decreases [10]. This inverted-U pattern suggested that disclosure-like behaviors aided trust when they reduced uncertainty in early relationship stages but became liabilities when they increased or persisted inappropriately. The benefit resided in the timing and moderation of the behavior rather than the behavior itself.

Temporal Dynamics

Trust development showed both immediate and cumulative effects. Single-session interactions could produce measurable trust and closeness gains [23, 24], particularly when structured disclosure facilitated reciprocal sharing [24]. However, these immediate effects were moderated by specific trust in the interaction partner [24], suggesting that rapid trust development required not just disclosure behaviors but also baseline willingness to trust.

Longer-term patterns revealed trust as fundamentally dynamic rather than simply increasing. The reciprocal cycle of concealment-distrust-further concealment operated at both weekly and daily timescales [9, 9], demonstrating that negative spirals could develop quickly. In contrast, positive trust development showed nonlinear growth with rapid initial phases followed by stabilization [1], suggesting different temporal signatures for trust building versus trust erosion.

The interplay between direct experience and prior information also showed temporal complexity. Indirect behavioral information influenced trust even after dozens of subsequent direct interactions [13], indicating that initial impressions established through disclosure or observation cast long shadows. This persistence of early information meant that the timing of disclosure mattered enormously—early disclosures shaped trajectories that later disclosures struggled to alter.

Neural mechanisms suggested staged trust development. The model proposed evolution from calculus-based trust (evaluating rewards/costs via ventral tegmental area) through knowledge-based trust (inferring intentions via paracingulate cortex) to identification-based trust (attachment via septal area) [19]. This neurological sequence paralleled behavioral findings of trust progressing from role-based to personal cue-based [3], providing converging evidence for qualitatively distinct trust phases requiring different disclosure types and depths.

References

1. Lisa van der Werff, Finian Buckley (2017) Getting to Know You. *Journal of Management*. <https://doi.org/10.1177/0149206314543475>
2. L. V. D. Werff (2018) TRUST DEVELOPMENT DURING SOCIALIZATION 1 Getting to know you: A longitudinal examination of trust cues and trust development during socialization
3. L. V. D. Werff (2014) On what bases? The role of trust cues in longitudinal trust development during newcomer socialisation

4. D. Bell (2018) Risk in New Sexual Relationships: Trajectories of Protection*. *Journal of Social Structure*. <https://doi.org/10.21307/joss-2018-008>
5. Qiong Chen, Qianqian Zhang, Si Zhao, Caina Li (2023) Trusting strangers: The benefits of reciprocal self-disclosure during online computer-mediated communication and mediating role of interpersonal liking. *International Journal of Psychology*. <https://doi.org/10.1002/ijop.12957>
6. Jusheng Liu, Jian-jia He, Shengxue He, et al (2022) Patients' Self-Disclosure Positively Influences the Establishment of Patients' Trust in Physicians: An Empirical Study of Computer-Mediated Communication in an Online Health Community. *Frontiers in Public Health*. <https://doi.org/10.3389/fpubh.2022.823692>
7. Corrie B. Whitmore, J. Dunsmore (2014) Trust Development: Testing a New Model in Undergraduate Roommate Relationships. *The Journal of Genetic Psychology*. <https://doi.org/10.1080/00221325.2013.869533>
8. John D. Davis (1976) Self-disclosure in an acquaintance exercise: Responsibility for level of intimacy. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/0022-3514.33.6.787>
9. Ahmet Uysal, H. L. Lin, A. L. Bush (2012) The reciprocal cycle of self-concealment and trust in romantic relationships. *European Journal of Social Psychology*. <https://doi.org/10.1002/EJSP.1904>
10. Qiyang Du (2025) Drink With Me: Alcohol's Dynamic Effect on Developing Supervisory Trust in Newcomers. *Academy of Management Proceedings*. <https://doi.org/10.5465/amproc.2025.13436abstract>
11. Young-ok Yum, K. Hara (2005) Computer-Mediated Relationship Development: A Cross-Cultural Comparison. *J Comput Mediat Commun*. <https://doi.org/10.1111/J.1083-6101.2006.TB00307.X>
12. Fanzheng Yang, Tanya Rosenblat, Weiwei Weng (2025) Managing trust under dynamic information. *Humanities and Social Sciences Communications*. <https://doi.org/10.1057/s41599-025-06248-y>
13. Pareezad Zarolia, Max Weisbuch, Kateri McRae (2017) Influence of Indirect Information on Interpersonal Trust Despite Direct Information. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/pspi0000074>
14. J. Ng, Oluremi B. Ayoko (2011) Conflict and Trust: A Study of the Dynamics between Intra-Team Conflict and Trust Over Time. <https://doi.org/10.2139/SSRN.1872608>
15. C. A. Vanlear (1987) The formation of social relationships. A longitudinal study of social penetration. *Human Communication Research*. <https://doi.org/10.1111/J.1468-2958.1987.TB00107.X>
16. M. Serva, Mark A. Fuller, R. Mayer (2005) The reciprocal nature of trust: A longitudinal study of interacting teams. *Journal of Organizational Behavior*. <https://doi.org/10.1002/JOB.331>
17. S. Altschuller, R. Benbunan-Fich (2008) In search of trust for newly formed virtual disaster recovery teams. *International Journal of Technology, Policy and Management*. <https://doi.org/10.1504/IJTPM.2008.020165>

18. S. Murray, Maya Aloni, J. G. Holmes, et al (2009) Fostering partner dependence as trust insurance: the implicit contingencies of the exchange script in close relationships. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/a0012856>
19. F. Krueger, K. McCabe, J. Moll, et al (2007) Neural correlates of trust. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.0710103104>
20. H. A. Mcallister (1980) Self-disclosure and liking: Effects for senders and receivers. *Journal of Personality*. <https://doi.org/10.1111/J.1467-6494.1980.TB02376.X>
21. S. Murray, J. G. Holmes, Jaye L. Derrick, et al (2013) Cautious to a Fault: Self-Protection and the Trajectory of Marital Satisfaction. *Journal of Experimental Social Psychology*. <https://doi.org/10.1016/J.JESP.2012.10.010>
22. Kari. A. Terzino, S. Cross (2009) Predicting Commitment in New Relationships: Interactive Effects of Relational Self-construal and Power. *Self and Identity*. <https://doi.org/10.1080/15298860802102273>
23. Matthew K. Miller, R. Mandryk (2021) Meeting with Media: Comparing Synchronous Media Sharing and Icebreaker Questions in Initial Interactions via Video Chat. *Proc ACM Hum Comput Interact*. <https://doi.org/10.1145/3479518>
24. T. Schneeberger, Anna Lea Reinwarth, Robin Wensky, et al (2023) Fast Friends: Generating Interpersonal Closeness between Humans and Socially Interactive Agents. *International Conference on Intelligent Virtual Agents*. <https://doi.org/10.1145/3570945.3607302>
25. R. Larzelere, T. Huston (1980) The Dyadic Trust Scale: Toward Understanding Interpersonal Trust in Close Relationships. *Journal of Marriage and Family*. <https://doi.org/10.2307/351903>