



INSTITUTE FOR HUMAN EVOLUTION  
IHE WHITEPAPER SERIES

---

IHE-WP-05

# Seductics: Toward a Probabilistic Framework for Romantic Outcomes

*Assumptions, Model Classes, and Validation Requirements*

---

|                    |                                                        |
|--------------------|--------------------------------------------------------|
| PUBLICATION NUMBER | IHE-WP-05                                              |
| SERIES             | IHE Whitepaper Series                                  |
| VERSION            | 1.0 (October 2026)                                     |
| DOCUMENT TYPE      | Whitepaper. Not peer reviewed.                         |
| PUBLISHER          | Institute for Human Evolution, Los Angeles, California |

## SUGGESTED CITATION

Institute for Human Evolution. (2026). *Seductics: Toward a Probabilistic Framework for Romantic Outcomes: Assumptions, Model Classes, and Validation Requirements* (IHE Whitepaper No. 5). Institute for Human Evolution. <https://ihe.institute/research#whitepapers>

## Abstract

**Purpose.** This paper sets out *Seductics*, the Institute's term for the probabilistic study of romantic outcomes, as a modeling framework whose assumptions can be tested rather than as a predictor. **Approach.** The paper reviews prior formal models of romantic dynamics, the record of statistical prediction in psychology, and the empirical literature on predicting initial attraction, relationship quality and dissolution. It then analyzes a naive baseline that treats interactions as independent trials, along with composite scoring indices and discrete trajectory typologies, and sets out alternative model classes: discrete-time Markov chains with absorbing states, hidden Markov models, discrete-time hazard models, Beta-Binomial updating, and the social relations model for dyadic data. Worked examples, computed in code with illustrative (not empirical) parameters, show how uncertainty should be reported. **Position.** The baseline  $P = 1 - (1 - p)^n$  assumes independent, identically distributed trials and fails for repeated interaction within a dyad; with heterogeneous or declining per-interaction probabilities it overstates cumulative probability, in one example by roughly a factor of two. Published evidence indicates that relationship-specific initial desire is difficult to predict before two people meet, and that change in relationship quality is largely unpredictable from self-report. **Limits.** The framework is suited to population-level description, simulation, power analysis and study design. IHE offers no individual-level predictions and no consumer tool. The paper closes with a staged validation roadmap that specifies the results under which the framework, or parts of it, should be abandoned.

**Keywords:** *probabilistic modeling; romantic attraction; relationship dissolution; Markov chains; survival analysis; Bayesian inference; social relations model; prediction; validation*

---

## CONTENTS

### 1. Introduction

- 1.1 Purpose and scope
- 1.2 Definitions

### 2. Prior formal work and the predictive record

- 2.1 Formal models of romantic dynamics already exist
- 2.2 Statistical versus intuitive prediction
- 2.3 What predicts initial attraction
- 2.4 What predicts relationship quality and dissolution
- 2.5 Divorce prediction and the problem of cross-validation
- 2.6 The general limits of predicting life outcomes

### 3. A naive baseline: interactions as independent trials

- 3.1 What the baseline computes
- 3.2 Assumption 1: a well-defined event
- 3.3 Assumption 2: independence and identical distribution
- 3.4 Worked example: heterogeneity and decay
- 3.5 Repetition, contact and timing

### 4. Composite indices and trajectory typologies

- 4.1 Composite indices as measurement problems
- 4.2 Attachment
- 4.3 Named trajectory curves

### 5. Candidate model classes

- 5.1 Discrete-time Markov chains with absorbing states
- 5.2 Hidden Markov models for latent interest
- 5.3 Hazard and survival models for time to dissolution
- 5.4 Bayesian updating with explicit priors
- 5.5 Dyadic and multilevel structure: the social relations model
- 5.6 Dynamical systems and summary

## **6. Validation requirements**

- 6.1 Identifiability
- 6.2 Measurement validity
- 6.3 Discrimination, calibration and proper scoring
- 6.4 Out-of-sample validation and overfitting
- 6.5 Base rates
- 6.6 Pre-registration, transparency and reporting
- 6.7 Fairness and generalizability across groups
- 6.8 Reactivity and Goodhart's law

## **7. Evidence at a glance**

## **8. A staged validation roadmap**

## **9. Research uses**

## **10. Ethical considerations**

## **11. Limitations and open questions**

## **12. Research agenda**

## **13. Conclusion**

## **References**

# 1. Introduction

---

## 1.1 Purpose and scope

People have long described courtship in the language of chance: a meeting is "lucky", a relationship "had good odds", a breakup was "bound to happen". Relationship science has turned parts of this intuition into testable models of who is attracted to whom at first meeting, which couples remain satisfied, and when relationships end. The Institute for Human Evolution (IHE) uses the term *Seductics* for the mathematical study of romantic probability, organized around three questions: how likely is attraction to form, how likely is it to deepen into attachment, and how likely is attachment to stabilize or end.

These questions invite simple answers that do not survive scrutiny. Popular and applied writing on dating often treats each interaction as a fresh roll of the dice, scores people on composite indices of "attraction potential" or "collapse risk", sorts relationships into a few named trajectory types, and offers timing rules for steering another person's interest. This paper treats the probabilistic view as a *proposed modeling framework*. It asks what assumptions such a framework requires, which model classes fit the structure of romantic interaction, and what evidence would be needed before any part of it could be trusted.

The paper is methodological and reports no new data. It is the fifth paper in the IHE series: IHE-WP-01 sets out the series' multilevel framework and definitions, IHE-WP-02 reviews the neurobiology of relationship loss, IHE-WP-03 the evolutionary literature on mating, and IHE-WP-04 the evidence on human chemosignals.

## 1.2 Definitions

Following IHE-WP-01, *seduction* is used in this series as an umbrella term for consensual, communicative behavior by which one person seeks to elicit or increase another person's romantic or sexual interest. Coercion, deception about material facts, harassment, and attempts to manipulate a person who has not consented are outside this definition. They are not treated as skills to be modeled or improved, and no model in this paper has the success of such behavior as its outcome.

A *romantic outcome* is a reportable state of a relationship between two people, defined in advance: for example, both people reporting interest in meeting again, both describing themselves as in a relationship, or the relationship having ended. Outcomes are properties of the *dyad* (the pair), established by both people's reports or by records they have consented to share, and never defined as one person's success in obtaining something from the other.

A *model* is a formal statement of how observable quantities are generated, including which are random and how they depend on one another; a *parameter* is a fixed but unknown quantity in a model. *Prediction* means stating a probability for an outcome before it is observed; *explanation* means identifying its causes, and good prediction does not imply correct explanation (Yarkoni & Westfall, 2017). *Population-level* statements concern rates in a defined population; *individual-level* statements

concern one person or couple. A model can describe a population well and still say little about any particular pair.

*Calibration* is the agreement between predicted probabilities and observed frequencies: among cases given a probability of 0.30, about 30% should experience the outcome. *Discrimination* is the ability to rank cases that experience the outcome above those that do not, commonly summarized by the area under the receiver operating characteristic curve (AUC), where 0.5 is chance and 1.0 perfect. *Out-of-sample validation* means evaluating a model on data not used to build it.

### Evidence-strength legend used in this paper

**Established:** replicated, convergent methods, large or meta-analytic evidence. **Supported:** consistent findings, but limited replication, samples, or methods. **Preliminary:** few or small studies, or a single method. **Contested:** credible evidence on more than one side, or failed replications. **Speculative:** plausible hypothesis, no direct evidence.

### What this paper does not claim

This paper does not offer a tool that predicts whether a particular person will become attracted to, form a relationship with, or leave a particular other person; IHE offers no individual-level predictions and no consumer tool. It does not claim that any index, score or equation described here has been validated, or that romantic behavior follows laws: the Institute describes tendencies, and every numerical example uses illustrative parameters that are not estimates from data. It does not claim that IHE has run studies on this topic or holds data on it. IHE provides preventive education, not treatment, and does not offer matchmaking. Nothing here is advice on how to influence another person's feelings, and people who need clinical care should see a licensed clinician.

## 2. Prior formal work and the predictive record

### 2.1 Formal models of romantic dynamics already exist

Romantic interaction has been modeled formally many times. Strogatz (1988) used a pair of linear differential equations, in a deliberately playful one-page note, to show how two people's feelings could cycle, converge or diverge depending on how each responds to the other. Felmlee and Greenberg (1999) proposed a dynamic systems model in which each partner's behavior adjusts toward a personal ideal and toward the partner's behavior. Gottman et al. (1999) fitted nonlinear difference equations to newlywed conflict interaction, estimating "influence functions" that describe how one spouse's affect shifts the other's. Butler (2011) reviewed work on temporal interpersonal emotion systems, mapping coupling, synchrony and co-regulation onto statistical models, and methods for non-independent dyadic data are well developed (Kenny et al., 2006). Any new framework must therefore say what it adds (**Established**, for the existence of such models). Most of these models are theoretical or fitted to modest samples: they show that formalization is possible, not that it yields accurate individual-level prediction.

## 2.2 Statistical versus intuitive prediction

Meehl (1954) reviewed early comparisons of clinical and statistical prediction and found statistical combination of information generally as accurate as clinical judgment or more so. Dawes (1979) showed that "improper" linear models, such as unweighted sums of standardized predictors, often outperform intuitive judgment and approach optimally weighted models in new samples. A meta-analysis by Grove et al. (2000) found mechanical prediction modestly more accurate than clinical prediction on average across many domains, and substantially less accurate in only a small minority of comparisons (**Established**). This supports formal modeling in general but does not show that romantic outcomes are highly predictable. It shows that when information is weakly predictive, simple formal combination extracts what signal there is more consistently than intuition, and that the validity of a scoring rule's components matters more than its weights.

## 2.3 What predicts initial attraction

Speed-dating designs, in which each participant has brief meetings with many potential partners, allow initial attraction to be studied with repeated measures on real interactions (Finkel et al., 2012). Each rating of romantic desire can be decomposed using the *social relations model* (SRM) into variance due to the rater's general tendency to desire others (actor variance), the ratee's general tendency to be desired (partner variance), and the specific pairing (relationship variance) (Kenny & La Voie, 1984; Kenny et al., 2006).

Joel et al. (2017) asked whether more than 100 self-reported traits and preferences, measured before two speed-dating events, could predict these components using random-forest models. The models predicted 4% to 18% of actor variance and 7% to 27% of partner variance, but could not predict relationship variance from any combination of pre-date measures. Who tends to like others and who tends to be liked were partly predictable; who would like *whom* was not (**Supported**: two studies, heterosexual speed-dating samples). Consensus about who is desirable is substantial among people who have just met but less marked as acquaintance develops, when relationship-specific evaluations become more prominent (Eastwick & Hunt, 2014). Stated ideal-partner preferences also predict evaluations of real partners less well than their prominence in self-report suggests, and sex differences in stated preferences are not well mirrored in how traits predict evaluations of actual partners (Eastwick et al., 2014).

Round-robin studies confirm that actor, partner and relationship components can be separated at zero acquaintance (Back et al., 2011). A one-year follow-up of a community speed-dating sample (382 participants aged 18 to 54) found that the chance of later sexual contact with a given speed-dating partner was 6% and of a romantic relationship 4% (Asendorpf et al., 2011), figures useful as base rates for study design though specific to one population and procedure. One further finding bears directly on Section 3: participants who expressed desire for many partners indiscriminately tended to be desired less in return, whereas selective desire tended to be reciprocated (Eastwick et al., 2007). The meaning of a signal depends on its pattern across partners, not only on its content.

## 2.4 What predicts relationship quality and dissolution

A meta-analysis of non-marital dissolution found that relationship-level variables such as commitment, love and dependence predicted breakup more strongly than individual-level variables such as personality (Le et al., 2010). A review of longitudinal research on marriage found many individual predictors with small effects and proposed a vulnerability-stress-adaptation model in which enduring vulnerabilities, stressful events and adaptive processes jointly shape trajectories (Karney & Bradbury, 1995).

The largest test to date is the machine-learning analysis by Joel et al. (2020) of 43 longitudinal datasets comprising 11,196 couples. People's own perceptions of the relationship, such as how satisfied and committed they judged their partner to be and how appreciative they felt, accounted for approximately 45% of the variance in their current relationship quality. The partner's reports added little, nor did either person's individual differences, and no variable predicted whose relationship quality would increase or decrease over time (**Supported**). These concurrent associations are not causal and partly reflect shared method, since predictors and outcomes were self-reported by the same person.

## 2.5 Divorce prediction and the problem of cross-validation

Gottman and Levenson (1992) reported that laboratory observations of interaction and physiology were associated with later marital dissolution, and Gottman et al. (1998) that newlywed interaction predicted later happiness and stability. A 14-year follow-up reported that a model including marital satisfaction, thoughts of dissolution and affective interaction predicted divorce with 93% accuracy (Gottman & Levenson, 2000). These findings have been sharply criticized. Stanley et al. (2000) raised concerns about non-random sampling, failure to control for initial differences, statistical ambiguities, and the use of correlational data to recommend changes to therapy. Heyman and Slep (2001) reanalyzed archival data and showed that accuracy and predictive value fell precipitously under cross-validation and once the population prevalence of divorce was considered. Observed interaction carries some prospective information about relationship outcomes (**Supported**); very high individual-level accuracy has not been demonstrated in independent samples (**Contested**). Section 6.4 shows why the distinction matters.

## 2.6 The general limits of predicting life outcomes

In a mass collaboration, 160 teams built models to predict six life outcomes from a rich birth-cohort dataset. The best predictions were not very accurate and only slightly better than a simple benchmark, and error depended far more on which family was predicted than on the technique used (Salganik et al., 2020). Romantic outcomes depend on two people, changing circumstances and chance encounters, and there is little reason to expect them to be more predictable. Claims of accurate individual-level romantic prediction should be treated as extraordinary, and every model should be compared with a simple benchmark such as the base rate.

### 3. A naive baseline: interactions as independent trials

#### 3.1 What the baseline computes

A naive baseline treats courtship as a sequence of independent attempts and computes

$$P(n) = 1 - (1 - p)^n \quad (1)$$

where  $p$  is the probability of a positive outcome in a single interaction,  $n$  is the number of interactions, and  $P(n)$  is the probability of at least one positive outcome in  $n$  interactions. Equation (1) is correct for  $n$  independent Bernoulli trials (trials with two outcomes) sharing the same  $p$ : every trial fails with probability  $(1 - p)^n$ , and the complement is at least one success. With  $p = 0.10$ ,  $P$  reaches 0.5 after seven trials and 0.651 after ten. Equation (1) does not compute the probability that attraction "accumulates" or "deepens"; it computes the probability that a binary event occurs at least once. Its application to courtship requires three assumptions, each of which fails.

#### 3.2 Assumption 1: a well-defined event

Accounts of this kind often leave "success" undefined. If success means one person's goal being met regardless of the other's wishes, the model falls outside the Institute's definition of seduction (Section 1.2) and its ethics (Section 10). Defined dyadically and consensually, for example as both people reporting interest in further contact, it becomes a measurable research outcome. The definition must be fixed in advance, because changing it after seeing data inflates apparent accuracy (Simmons et al., 2011).

#### 3.3 Assumption 2: independence and identical distribution

Equation (1) treats each interaction as a fresh draw. Between the same two people this is implausible. Each exchange changes what each person knows and feels, so the probability of a positive response depends on earlier steps (history dependence). Interest is often mutual and responsive to the other's interest (reciprocity), though not uniformly (Eastwick et al., 2007). People learn, tire, and experience changing circumstances (non-stationarity). Dyadic data are non-independent by construction, and analyses that ignore this mislead (Kenny et al., 2006; Butler, 2011).

Equation (1) is more defensible for attempts with *different* partners, such as a series of first meetings. Even there,  $p$  varies across pairs, and a person's prospects in different encounters are correlated through stable partner variance (Section 2.3). In online dating, contact is structured by a shared desirability hierarchy, and users often contact people ranked above themselves, from whom replies are less likely (Bruch & Newman, 2018).

#### 3.4 Worked example: heterogeneity and decay

Two departures from Equation (1) show the size of the error. First, let  $p$  vary across pairs according to a Beta distribution with mean 0.10. Because  $(1 - p)^n$  is convex in  $p$ , averaging over heterogeneous pairs gives a lower cumulative probability than inserting the average  $p$  into Equation (1) (Jensen's inequality):

$$P(n) = 1 - E[(1 - p)^n] = 1 - B(a, b + n) / B(a, b) \quad (2)$$

where  $p \sim \text{Beta}(a, b)$ ,  $E$  is expectation over pairs, and  $B$  is the Beta function. Second, let the per-interaction probability within one dyad decline with repetition,  $p_t = p_0 \delta^{t-1}$ , with  $p_0 = 0.10$  and decay factor  $\delta = 0.85$ :

$$P(n) = 1 - \prod_{t=1}^n (1 - p_t) \quad (3)$$

**Table 1.** Probability of at least one positive outcome after  $n$  interactions under four assumptions (illustrative parameters computed in code; per-interaction probability 0.10 on average at the first interaction in every column).

| $n$ | Constant $p$ (Eq. 1) | $p \sim \text{Beta}(5, 45)$ | $p \sim \text{Beta}(0.5, 4.5)$ | Declining $p_t$ (Eq. 3) |
|-----|----------------------|-----------------------------|--------------------------------|-------------------------|
| 1   | 0.100                | 0.100                       | 0.100                          | 0.100                   |
| 5   | 0.410                | 0.397                       | 0.322                          | 0.320                   |
| 10  | 0.651                | 0.619                       | 0.453                          | 0.425                   |
| 20  | 0.878                | 0.830                       | 0.581                          | 0.483                   |
| 40  | 0.985                | 0.954                       | 0.690                          | 0.496                   |

With modest heterogeneity, Equation (1) is close to the heterogeneous result. With strong heterogeneity ( $\text{Beta}(0.5, 4.5)$ , where many pairs have near-zero probability and a few high probability), it overstates  $P(20)$  by about 0.30. Under decay, the infinite product in Equation (3) converges, so  $P(n)$  never exceeds 0.496 however many interactions occur; at  $n = 40$  Equation (1) gives 0.985, roughly double. The conclusion does not depend on the particular numbers: whenever per-interaction probabilities vary across pairs or fall with repetition, Equation (1) overstates cumulative probability, increasingly so as  $n$  grows.

### 3.5 Repetition, contact and timing

A common elaboration holds that more interaction raises cumulative probability while  $p$  stays positive, but that once  $p$  falls to zero further contact becomes counterproductive, explaining why excessive messaging or persistent pursuit fails. The argument is internally inconsistent. Under Equation (1), if  $p = 0$  then  $P$  stays constant as  $n$  grows; it cannot fall. Counterproductive contact requires  $p$  to depend on the history of interaction, which Equation (1) excludes. The intuition needs a model with feedback (Section 5).

Whether more exposure increases liking is itself nuanced. Mere exposure reliably increases liking on average, with boundary conditions including the number and duration of exposures and the type of stimulus (Bornstein, 1989). For people, Norton et al. (2007) reported that learning more about someone tended on average to reduce liking as dissimilarities emerged, while Reis et al. (2011) reported that greater familiarity through live interaction increased attraction. The contact-attraction relationship is **Contested** and likely depends on what contact reveals.

Timing heuristics (infrequent contact loses momentum, frequent contact lowers perceived value, strategic spacing raises interest) are rarely accompanied by evidence. In one experiment, women

viewing profiles of men whose interest was uncertain were more attracted, and thought about the men more, than when the men's interest was known to be high (Whitchurch et al., 2011). This single study, in a profile-viewing paradigm with one gender configuration, does not test message spacing and concerns a short-lived state rather than a relationship outcome (**Preliminary** for the uncertainty effect; **Speculative** for timing heuristics). Such heuristics are also framed as tactics for steering another person's feelings, which the Institute does not pursue (Section 10); they appear here only as hypotheses for ethically constrained tests (Section 12).

## 4. Composite indices and trajectory typologies

---

### 4.1 Composite indices as measurement problems

A tempting step is to define composite indices, such as "attraction potential", "emotional responsiveness", "stability projection" or "collapse vulnerability", score each from 0 to 1, and combine them into a forecast. Any such index must first meet ordinary measurement standards: items, scoring rules, reliability and validation data. Construct validity requires a theory linking the construct to observable indicators and to other constructs, and evidence that the measure behaves as predicted (Cronbach & Meehl, 1955). Many published measures lack such evidence, and undisclosed measurement choices are a major source of unreliable findings (Flake & Fried, 2020).

Such indices have specific problems. An index of "collapse vulnerability" is defined by the outcome it predicts and is circular unless measured strictly beforehand. Constructs such as "microexpression chemistry" have no operational definition. Trauma history is sensitive clinical information that a relationship model should not score. Rescaling disparate quantities to 0 to 1 does not make them commensurable. Where constructs overlap with established ones (commitment, attachment, responsiveness), incremental validity must be shown, and Section 2.4 suggests that well-measured relationship perceptions already capture much of what self-report offers (Joel et al., 2020). If an index is ever built from valid components, equal weighting is a sensible default that resists overfitting (Dawes, 1979). And once a score becomes a target, behavior changes in ways that can destroy its meaning, a general form of Goodhart's law (Strathern, 1997) (Section 6.8).

### 4.2 Attachment

Applied accounts often assign fixed profiles to "secure", "avoidant" and "anxious" individuals. Hazan and Shaver (1987) showed that adults' romantic experiences could be organized in attachment terms, but later work generally measures adult attachment on two continuous dimensions (anxiety and avoidance) and finds attachment security only moderately stable over long periods (Fraley, 2002). Attachment orientations also vary across relationships and change with experience. Assigning types and matching score profiles discards this information and invites stereotyping. In this framework, attachment dimensions are candidate covariates whose contribution must be estimated (**Speculative** for fixed type-to-profile mappings).

### 4.3 Named trajectory curves

Relationships are often described as following a few named curves, such as steady ascent, rapid rise then volatility, push-and-pull oscillation, or erratic change. Whether trajectories fall into discrete classes is an empirical question with treacherous methods. Group-based trajectory models approximate continuous variation with a finite number of groups, best understood as approximations rather than literal types (Nagin, 1999), and growth mixture models can extract spurious classes from merely non-normal data (Bauer & Curran, 2003). Such typologies are **Speculative**. A defensible test would pre-register class enumeration, compare discrete and continuous models on held-out data, and require independent replication (Section 12).

## 5. Candidate model classes

The following standard model classes match the structure of romantic interaction better than Equation (1). For each, the paper states the question it addresses, what it assumes, and how uncertainty should be reported. All numerical examples are illustrative, computed in code, and are not estimates of real-world quantities.

### 5.1 Discrete-time Markov chains with absorbing states

A discrete-time Markov chain occupies one of a finite set of states at each step and moves between them with fixed probabilities that depend only on the current state (Norris, 1997). With  $\pi_t$  the row vector of state probabilities at step  $t$  and  $P$  the transition matrix,

$$\pi_{t+1} = \pi_t P \quad (4)$$

where  $P_{ij}$  is the probability of moving from state  $i$  to state  $j$  and each row sums to 1. An *absorbing* state, once entered, is never left. Partitioning  $P$  into  $Q$  (transitions among transient states) and  $R$  (transient to absorbing), the *fundamental matrix*

$$N = (I - Q)^{-1} \quad (5)$$

gives the expected number of steps in each transient state before absorption, where  $I$  is the identity matrix. Absorption probabilities are  $B = NR$ , and expected steps to absorption are  $t = N\mathbf{1}$ , where  $\mathbf{1}$  is a column vector of ones.

Consider three transient states, initial contact (I), mutual interest (M) and ongoing relationship (R), and two absorbing states: contact or relationship ended (E), and a pre-specified commitment milestone reached (C), such as both partners reporting a committed relationship at follow-up. Absorption is a design feature: a relationship reaching C can later end, but the model concerns first passage to C or E within the study.

**Table 2.** Illustrative monthly transition matrix (rows: current state; columns: next state). Values are invented for exposition.

| From / to           | I    | M    | R    | E    | C    |
|---------------------|------|------|------|------|------|
| Initial contact (I) | 0.30 | 0.25 | 0.00 | 0.45 | 0.00 |

| From / to                | I    | M    | R    | E    | C    |
|--------------------------|------|------|------|------|------|
| Mutual interest (M)      | 0.05 | 0.35 | 0.30 | 0.30 | 0.00 |
| Ongoing relationship (R) | 0.00 | 0.05 | 0.80 | 0.08 | 0.07 |

From initial contact, Equation (5) and  $B = NR$  give a probability of 0.067 of reaching C and 0.933 of reaching E, with an expected 3.08 months to absorption. From mutual interest the probability of reaching C is 0.188 (4.62 months), and from an ongoing relationship 0.397 (6.15 months).

These point values conceal parameter uncertainty, which propagates nonlinearly through Equation (5). Treating each row as Dirichlet-distributed around the values in Table 2, with concentration equivalent to about 40 observed transitions per state, 20,000 simulated draws give the probability of reaching C from initial contact a median of 0.060 and a 95% interval of 0.014 to 0.161; expected time to absorption has a 95% interval of 2.17 to 5.01 months. With the equivalent of 400 transitions per state, the interval for reaching C narrows to 0.045 to 0.093. A probability reported without such an interval is not a scientific estimate.

The chain's assumptions must be tested. The *Markov property* says the next state depends only on the current one; *time homogeneity* says transition probabilities do not change with time or relationship duration. Both can be checked against higher-order or duration-dependent alternatives on held-out data, and heterogeneity can enter through covariates or random effects. Even a well-fitting chain describes a population; a single couple's path remains highly uncertain.

## 5.2 Hidden Markov models for latent interest

Interest is not directly observed; researchers observe behaviors and reports such as reply latency, ratings, or whether a second meeting occurs. A *hidden Markov model* (HMM) treats the underlying state as latent (Visser, 2011):

$$P(z_t = j \mid z_{t-1} = i) = A_{ij}, \quad P(y_t \mid z_t = j) = f_j(y_t) \quad (6)$$

where  $z_t$  is the latent state at time  $t$  (for example, low, moderate or high mutual interest),  $A$  the latent transition matrix,  $y_t$  the observation, and  $f_j$  the distribution of observations in state  $j$ . HMMs separate noisy behavior from slower-changing states but carry three risks. The number of states may be poorly determined by data, with the over-extraction hazards of Section 4.3. State labels such as "high interest" come from the researcher and are hypotheses about construct validity, not findings. And parameters may be weakly identified when sequences are short (Section 6.1). Reading a behavior such as slow replies as evidence of someone's inner state is also an inference about that person, with the ethical implications of Section 10.

## 5.3 Hazard and survival models for time to dissolution

Whether a relationship stabilizes or ends is a question about *time to an event*. Survival analysis handles it, including relationships that have not ended by the study's end (right censoring) (Cox, 1972; Singer

& Willett, 1993). With  $T$  the month a relationship ends, the discrete-time hazard and survival function are

$$h(t) = P(T = t \mid T \geq t), \quad S(t) = \prod_{k=1}^t (1 - h(k)) \quad (7)$$

and covariates commonly enter as logit  $h(t \mid x) = \alpha_t + \beta^T x$ , where  $\alpha_t$  is a baseline term for each period,  $x$  a covariate vector and  $\beta$  its coefficients (Singer & Willett, 1993); the continuous-time counterpart is the proportional hazards model (Cox, 1972). Table 3 compares an illustrative declining hazard (0.06 per month in months 1 to 6, 0.035 in months 7 to 12, 0.02 in months 13 to 24) with a constant hazard of 0.0339, chosen in code to give the same 24-month survival.

**Table 3.** Illustrative survival probabilities under a declining and a constant monthly hazard with equal 24-month survival.

| Month $t$ | $S(t)$ , declining hazard | $S(t)$ , constant hazard |
|-----------|---------------------------|--------------------------|
| 6         | 0.690                     | 0.813                    |
| 12        | 0.557                     | 0.661                    |
| 24        | 0.437                     | 0.437                    |

The models agree at 24 months but the constant hazard overstates six-month survival by more than 0.12. A single summary probability of "collapse" hides this, so dissolution models must report the hazard over time, whose shape is itself an empirical question. The same machinery can describe recovery after loss (IHE-WP-02), with the event defined as return to a pre-specified level of well-being.

## 5.4 Bayesian updating with explicit priors

Bayesian inference represents uncertainty about a parameter as a probability distribution updated by data (Gelman et al., 2013). For a proportion  $\theta$ , such as the share of first meetings in a study population after which both people request a second meeting,

$$\theta \sim \text{Beta}(a, b), \quad x \mid \theta \sim \text{Binomial}(n, \theta), \quad \theta \mid x \sim \text{Beta}(a + x, b + n - x) \quad (8)$$

where  $a$  and  $b$  are prior parameters,  $n$  the number of meetings and  $x$  the number with a mutual request. Table 4 shows posteriors for an invented result of 18 in 60 under three priors.

**Table 4.** Posterior distributions for an illustrative proportion ( $x = 18$ ,  $n = 60$ ); 95% equal-tailed credible intervals computed in code.

| Prior                                | Posterior    | Posterior mean | 95% interval   |
|--------------------------------------|--------------|----------------|----------------|
| Beta(1, 1), uniform                  | Beta(19, 43) | 0.306          | 0.199 to 0.426 |
| Beta(2, 2), weakly informative       | Beta(20, 44) | 0.312          | 0.206 to 0.430 |
| Beta(10, 30), informative, mean 0.25 | Beta(28, 72) | 0.280          | 0.197 to 0.371 |

Priors must be stated and their influence reported; here a moderately informative prior shifts the mean by about 0.03. With 60 observations the interval remains wide. Predictions must include parameter

uncertainty: under the weakly informative prior, the posterior predictive distribution for mutual requests in the next 20 meetings has mean 6.25 and a 95% interval of 2 to 11, wider than the 2 to 10 obtained by treating the posterior mean as known.  $\theta$  is a population proportion. Reusing the arithmetic for one person's chances with another, updated after each interaction, inherits the problems of Section 3, because interactions within a dyad are not exchangeable draws from a fixed  $\theta$ . Hierarchical models, in which each dyad has its own parameter drawn from a population distribution, are the appropriate structure (Gelman et al., 2013), and the spread of that distribution (the heterogeneity of Table 1) is often the key quantity.

### 5.5 Dyadic and multilevel structure: the social relations model

The social relations model decomposes a rating  $y_{ijk}$  by person  $i$  of person  $j$  on occasion  $k$  as

$$y_{ijk} = \mu + a_i + b_j + g_{ij} + e_{ijk} \quad (9)$$

where  $\mu$  is the grand mean,  $a_i$  the actor effect ( $i$ 's general tendency to rate others highly),  $b_j$  the partner effect ( $j$ 's general tendency to be rated highly),  $g_{ij}$  the relationship effect ( $i$ 's unique response to  $j$ ), and  $e_{ijk}$  occasion-specific error (Kenny & La Voie, 1984; Kenny et al., 2006). The full model also estimates reciprocity correlations. Equation (9) shows that "the probability of attraction" is not one quantity: it mixes rater tendencies, target tendencies and the specific pairing, which have different predictors (Joel et al., 2017).

It also shows what designs can identify. In 2,000 simulated replications of a design in which 10 people each rate the same 10 others (true variance shares 0.20 actor, 0.30 partner, 0.30 relationship, 0.20 error), estimates with two measurements per pair were unbiased on average but varied widely: 95% of actor-variance estimates fell between 0.03 and 0.48, and of partner-variance estimates between 0.07 and 0.70. With one measurement per pair, relationship and error variance could not be separated; only their sum (about 0.50) was identified. Claims about relationship-specific attraction therefore require repeated measures or multiple indicators per pair and samples large enough for useful precision.

### 5.6 Dynamical systems and summary

Dynamical models of the kind reviewed in Section 2.1 (Strogatz, 1988; Felmlee & Greenberg, 1999; Gottman et al., 1999) represent continuous mutual influence rather than discrete states. They suit intensive repeated measures, such as moment-to-moment affect in conversation, but are prone to fitting noise when parameters are many and series short, making identifiability analysis essential (Butler, 2011). Table 5 summarizes the model classes.

**Table 5.** Summary of candidate model classes.

| Model class            | Question addressed                 | Key assumption to test            | Main risk                          |
|------------------------|------------------------------------|-----------------------------------|------------------------------------|
| Absorbing Markov chain | Which states are reached, and when | Markov property; time homogeneity | False precision without intervals  |
| Hidden Markov model    | Latent interest behind behavior    | Number and meaning of states      | Over-extraction; mislabeled states |

| Model class                  | Question addressed                    | Key assumption to test                  | Main risk                     |
|------------------------------|---------------------------------------|-----------------------------------------|-------------------------------|
| Discrete-time hazard         | Timing of dissolution or recovery     | Hazard shape; non-informative censoring | Summaries that hide timing    |
| Beta-Binomial (hierarchical) | Population rates with uncertainty     | Exchangeability within level            | Misuse as individual forecast |
| Social relations model       | Actor, partner, relationship variance | Design supports decomposition           | Underpowered round-robins     |
| Dynamical systems            | Mutual influence over time            | Functional form of coupling             | Overfitting short series      |

## 6. Validation requirements

A model is useful for science only if it can be wrong in a detectable way. Any Seductics model must meet the following requirements before its outputs are interpreted.

### 6.1 Identifiability

A parameter is *structurally identifiable* if different values produce different data distributions, and *practically identifiable* if available data estimate it with useful precision (Raue et al., 2009). Section 5.5 showed a structural non-identifiability; hidden-state and dynamical models often have practical ones, in which several parameter combinations fit equally well. Profile-likelihood or posterior analyses should be reported for every interpreted parameter, and unidentifiable parameters should not be given substantive names.

### 6.2 Measurement validity

Each input and outcome needs evidence of reliability, construct validity (Section 4.1) and *measurement invariance*, meaning that it measures the same thing across the groups compared. Self-report dominates the literature (Joel et al., 2020); behavioral traces such as message timing are not automatically more valid, because their meaning depends on context, platform and culture.

### 6.3 Discrimination, calibration and proper scoring

Probabilistic predictions should be evaluated with *proper scoring rules*, which are optimized in expectation only by reporting one's true probabilities, such as the Brier score (mean squared error of probability forecasts) and the logarithmic score (Gneiting & Raftery, 2007). Discrimination and calibration should both be reported: a model can rank cases well yet give systematically wrong probabilities (Van Calster et al., 2019). Calibration should be assessed overall and within subgroups.

### 6.4 Out-of-sample validation and overfitting

Models evaluated on the data used to fit them overstate accuracy, most of all when predictors are many relative to cases (Babyak, 2004; Yarkoni & Westfall, 2017). In 200 simulated replications, a linear model was fitted to 60 couples using 20 pure-noise predictors and a binary outcome with base rate 0.30, independent of all predictors. Mean in-sample AUC was 0.869 (95% of replications between

0.789 and 0.953), suggesting strong discrimination; on 500 new simulated couples it was 0.498 (0.435 to 0.553), which is chance. Small couples studies with many observational codes are exactly this setting, which is why cross-validation changes the picture in divorce prediction (Heyman & Slep, 2001). Validation should use held-out data, ideally from another time or site, and compare the model with a simple benchmark (Salganik et al., 2020).

## 6.5 Base rates

The value of a prediction depends on how common the outcome is, and people tend to neglect base rates when judging predictive evidence (Kahneman & Tversky, 1973). Table 6 shows, for a hypothetical classifier that correctly identifies 80% of relationships that will end within a year and 80% of those that will not, the share of positive predictions that are correct (*positive predictive value*, PPV) and of negative predictions that are correct (*negative predictive value*, NPV).

**Table 6.** Predictive values of an illustrative classifier with 80% sensitivity and 80% specificity.

| Base rate | PPV   | NPV   | Correct positives per 1,000 | False positives per 1,000 |
|-----------|-------|-------|-----------------------------|---------------------------|
| 0.05      | 0.174 | 0.987 | 40                          | 190                       |
| 0.20      | 0.500 | 0.941 | 160                         | 160                       |
| 0.50      | 0.800 | 0.800 | 400                         | 100                       |

At a 5% base rate, more than four of five relationships flagged as likely to end would not end. Accuracy claims must state the base rate of the population concerned, and base rates differ greatly across settings: a 4% chance of a relationship with a given speed-dating partner (Asendorpf et al., 2011) says nothing about dissolution among established couples.

## 6.6 Pre-registration, transparency and reporting

Large replication efforts found that a substantial share of published psychological findings do not replicate at their original strength (Open Science Collaboration, 2015), and analytic flexibility can make almost any result appear significant (Simmons et al., 2011). Pre-registration of hypotheses, outcomes, model specifications and metrics separates confirmatory tests from exploration (Nosek et al., 2018). Reporting guidelines for prediction models specify what must be disclosed about data, predictors, missing data, development and validation (Collins et al., 2015); the Institute intends to follow them, adapted as needed, in any prediction research.

## 6.7 Fairness and generalizability across groups

Prediction models can encode disparities. A widely used health algorithm that predicted care costs as a proxy for need systematically underestimated the needs of Black patients (Obermeyer et al., 2019), a failure arising from the choice of target variable rather than the technique. When base rates differ between groups, a calibrated score generally cannot equalize false-positive and false-negative rates, so fairness criteria must be chosen explicitly (Chouldechova, 2017). Most relevant samples are Western,

educated, industrialized, rich and democratic (Henrich et al., 2010), and several key attraction studies cited here involve heterosexual dyads only (Joel et al., 2017; Bruch & Newman, 2018). Longitudinal comparison of gay, lesbian and heterosexual couples has found largely similar processes predicting relationship quality and stability (Kurdek, 2004), but this cannot be assumed for every model. Models must be evaluated across gender identities, sexual orientations, cultures, ages and relationship structures, including non-monogamous relationships, and not applied outside the populations in which they were validated.

## 6.8 Reactivity and Goodhart's law

People who know they are measured, or learn which behaviors a model rewards, change their behavior; when a measure becomes a target, it tends to lose its value as a measure (Strathern, 1997). A published "communication index", for instance, would invite people to manage their message timing to raise their score, breaking any relation the index had to genuine interest. This is a further reason the Institute does not publish scoring tools for individuals and treats validated research measures as descriptions of populations, not levers.

## 7. Evidence at a glance

**Table 7.** Evidence at a glance.

| Claim                                                                             | Evidence strength | Representative sources                                                     | Principal limitation                |
|-----------------------------------------------------------------------------------|-------------------|----------------------------------------------------------------------------|-------------------------------------|
| Formal models of romantic dynamics are decades old                                | Established       | Strogatz (1988); Felmlee and Greenberg (1999); Gottman et al. (1999)       | Mostly theoretical or small samples |
| Statistical rules match or beat unaided judgment                                  | Established       | Meehl (1954); Dawes (1979); Grove et al. (2000)                            | Few romantic-domain tests           |
| Relationship-specific initial desire is hard to predict before meeting            | Supported         | Joel et al. (2017); Eastwick and Hunt (2014)                               | Heterosexual speed-dating samples   |
| Relationship perceptions predict current quality; change is largely unpredictable | Supported         | Joel et al. (2020); Le et al. (2010)                                       | Self-report; mainly WEIRD samples   |
| High divorce-prediction accuracy survives cross-validation                        | Contested         | Gottman and Levenson (2000); Heyman and Slep (2001); Stanley et al. (2000) | Few independent prospective tests   |
| Interactions within a dyad are not independent trials                             | Supported         | Eastwick et al. (2007); Kenny et al. (2006); Butler (2011)                 | Dependence varies by context        |
| More contact steadily increases attraction                                        | Contested         | Bornstein (1989); Norton et al. (2007); Reis et al. (2011)                 | Boundary conditions unresolved      |
| Strategic timing raises the probability of interest                               | Speculative       | Whitchurch et al. (2011)                                                   | No direct test of spacing           |

| Claim                                                     | Evidence strength | Representative sources                                                | Principal limitation                     |
|-----------------------------------------------------------|-------------------|-----------------------------------------------------------------------|------------------------------------------|
| Attachment types map onto fixed score profiles            | Speculative       | Hazan and Shaver (1987); Fraley (2002)                                | Attachment is dimensional and changeable |
| Trajectories form a few discrete curves                   | Speculative       | Nagin (1999); Bauer and Curran (2003)                                 | Spurious class extraction                |
| Actor, partner and relationship variance are separable    | Established       | Kenny and La Voie (1984); Back et al. (2011); Asendorpf et al. (2011) | Requires round-robin or repeated designs |
| Prediction models can create unfairness and privacy harms | Supported         | Obermeyer et al. (2019); Chouldechova (2017); Kosinski et al. (2013)  | Evidence mainly from other domains       |

## 8. A staged validation roadmap

The framework would be tested in stages, each with a failure condition fixed in advance. A component that fails is revised and retested in new data or abandoned, not carried forward on the strength of theory. Thresholds below illustrate the kind of criterion each pre-registration would fix.

1. **Stage 0: specification.** Define states, outcomes and covariates operationally, with dyadic, consent-based outcomes only; specify models, priors, metrics and benchmarks; pre-register. Failure: states cannot be coded reliably (for example, inter-rater agreement below a pre-specified Cohen's  $\kappa$  such as 0.70), requiring revised definitions before any modeling.
2. **Stage 1: measurement.** Establish reliability, construct validity, incremental validity and measurement invariance for every proposed index. Failure: no incremental validity over established constructs such as commitment or attachment dimensions, in which case the index is dropped.
3. **Stage 2: structural tests.** Test Markov order, time homogeneity, number of latent states, hazard shape and identifiability on longitudinal data. Failure: the independence model of Equation (1) predicts held-out sequences as well as dependence models, which would falsify this paper's central critique; or a proposed structure, such as a four-class typology, fails to replicate.
4. **Stage 3: prospective population-level prediction.** Predict outcomes in a new cohort with pre-registered models, compared on proper scoring rules against the base rate and a simple established-predictor model. Failure: no improvement over benchmark with an interval excluding zero, or poor calibration (for example, a calibration slope outside a pre-specified range); the model is then not used for predictive claims.
5. **Stage 4: external validation.** Repeat Stage 3 across sites, cultures, ages and relationship types, including same-sex and non-monogamous relationships, with subgroup calibration and

fairness reporting. Failure: material miscalibration in any subgroup, restricting the model's stated scope accordingly.

The roadmap ends with research instruments, not consumer applications. A model passing all five stages would still describe populations, and the Institute would publish it, with data and code where consent allows, for scientific use.

## 9. Research uses

---

The framework is intended for research uses only.

- **Simulation for hypothesis generation.** Simulation makes the consequences of assumptions visible, as Table 1 does for heterogeneity and decay, and can show where verbal theories make different predictions.
- **Power analysis and sample-size planning.** Simulation-based power analysis suits complex designs (Lakens & Caldwell, 2021). The SRM simulation of Section 5.5 shows that ten people per side estimates actor and partner variance very imprecisely.
- **Study design.** Models clarify which measurements identify which quantities, how long follow-up must be to estimate hazards, and what base rates imply for detectable effects.
- **Evidence synthesis.** Model classes give a common language for comparing studies with different outcomes and time scales, for example by converting reported dissolution rates into hazards.

The framework is not a basis for dating strategy, coaching, matchmaking or clinical assessment. IHE's applied work is preventive education about relationships, not treatment. Any future IHE study collecting relationship data will run under institutional review board supervision with partner universities and the Institute's Independent Ethics Committee, and participants reporting significant distress will be referred to licensed clinicians.

## 10. Ethical considerations

---

**Consent and autonomy.** A model of romantic outcomes describes two people, one of whom may not know a model is being applied. Outcomes must be defined dyadically and consensually, data about a person collected only with that person's informed consent, and no model built to increase one person's influence over another. An index that treats another person's emotional caution as "resistance" to be overcome by pacing, for example, is incompatible with this position.

**Manipulation and coercive optimization.** A model that predicts another person's responses can be used to search for inputs that produce a desired response, turning a description of a person into a tool for steering them; this becomes coercive when it exploits vulnerability, persistence or uncertainty. The

finding that uncertainty about another's interest can heighten attraction (Whitchurch et al., 2011) could be misread as license to create uncertainty deliberately. The Institute does not publish individual-level optimization tools, scoring rules for use on other people, or tactical recommendations derived from models.

**Privacy of relationship and biometric data.** Facebook Likes alone allowed accurate prediction of sexual orientation, political views and other sensitive attributes (Kosinski et al., 2013). Message histories, location data and physiological recordings from relationships are at least as sensitive and concern the partner as well as the participant. Technology is already used in domestic violence and stalking to monitor and control partners (Woodlock, 2017), and relationship-prediction tools would extend that repertoire. The framework therefore excludes hormonal, chemosignal and physiological inputs, which lack a validated basis (see IHE-WP-02 and IHE-WP-04) and create acute privacy risks. Any physiological measurement in future research would require ethical review, data minimization, secure storage and the consent of each person measured.

**Dignity and the language of prediction.** Probabilities about relationships can be heard as verdicts. Telling someone their relationship has a high "collapse risk" may be inaccurate (Section 6.5), self-fulfilling and distressing. Population-level findings will be communicated as such, with uncertainty, and never translated into individual risk labels.

**Fairness and cultural sensitivity.** Courtship norms, the meaning of signals and relationship forms vary across cultures, communities and individuals, and a model trained in one setting can treat its norms as universal (Section 6.7). The framework must not assume heterosexual, monogamous, reproduction-focused or gender-binary relationships, and must state the populations its evidence covers.

## 11. Limitations and open questions

---

This paper reports no data, and none of the numbers in Tables 1 to 4 or 6 estimates a real-world quantity. The evidence on initial attraction comes largely from speed-dating and online-dating samples of young, heterosexual, Western participants, and online-dating evidence reflects platform designs that shape behavior. Evidence on relationship quality relies heavily on self-report, and concurrent associations are not causal. Divorce-prediction studies have been criticized for sampling, analysis and lack of cross-validation.

Open questions bear directly on the framework. How strong and variable is history dependence within dyads, and does a low-order Markov model capture it? How much of the unpredictability of relationship-specific attraction reflects measurement limits rather than genuine contingency, and would measures of the interaction itself change the picture? Are discrete states such as contact, mutual interest and relationship meaningful for relationship forms that do not follow conventional stages? How should chance encounters and external events be represented, given that they are sources of variance no

individual-level predictor can remove? And can predictive research here avoid creating data or tools usable to monitor or manipulate partners?

Prediction is also not the main scientific goal. A well-calibrated population model would say little about why relationships form or end; explanation requires experimental, longitudinal and qualitative work that formal models can organize but not replace, consistent with the multilevel approach of IHE-WP-01.

## 12. Research agenda

---

The following questions are aspirations for a young institute, framed so that they could be pre-registered and could fail.

1. **Dependence within dyads.** In consented longitudinal interaction records between new acquaintances, a first-order Markov model of positive responses will outperform the independent-trials model (Equation 1) in held-out log score. Failure would undercut this paper's central critique and would be reported.
2. **Heterogeneity across partners.** Counts of mutual interest across first meetings with different partners will be overdispersed relative to a binomial model, with a Beta-Binomial model fitting better out of sample.
3. **Generalization of the relationship-variance finding.** The unpredictability of relationship-specific desire from pre-meeting self-reports (Joel et al., 2017) will replicate in samples including same-sex attraction and in non-Western settings; successful prediction would be an important departure.
4. **Trajectory classes.** Cross-validated class enumeration in longitudinal relationship-quality data will not support a small set of stable named classes, and continuous growth models will fit as well or better in an independent sample (linking to IHE-WP-01 and IHE-WP-02).
5. **Contact and attraction.** In an ethically constrained design using profiles or structured conversations with full debriefing, the effect of information exchanged on attraction will depend on the similarity revealed, reconciling Norton et al. (2007) and Reis et al. (2011). The same design could test the inverted-U effect of contact spacing that popular timing heuristics imply.
6. **Hazard shape.** In existing longitudinal couples data, the dissolution hazard will vary with relationship duration, and constant-hazard models will be detectably miscalibrated in the first year.
7. **Stated preferences and outcomes.** Evolutionary and social-structural accounts of mate preference (IHE-WP-03) predict different pre-meeting carriers of partner variance, testable within the SRM framework.

Each study would be pre-registered, ethically reviewed, and published with null results reported as fully as positive ones.

## 13. Conclusion

---

The intuition behind Seductics, that romantic outcomes have regularities describable in probabilistic terms, is sound and has a substantial scientific history. Its simplest formalization is not. Treating interactions as independent, identical trials ignores the nature of interaction within a couple and overstates probabilities whenever chances vary across pairs or fall with repetition. Composite indices without measurement, typologies without evidence and timing rules without tests add the appearance of precision without its substance.

What remains is a set of established model classes (absorbing Markov chains, hidden Markov models, hazard models, hierarchical Bayesian updating and the social relations model) matched to the questions of attraction, deepening and dissolution, with explicit requirements for identifiability, measurement validity, calibration, out-of-sample testing against benchmarks, attention to base rates, pre-registration and fairness. The empirical record suggests that, even under these standards, individual-level prediction of romantic outcomes will remain weak; that is one of the framework's findings, not its failure. IHE offers no individual-level predictions and no consumer tool. Its interest is in how relationships tend to unfold across populations, stated with honest uncertainty, so that people can think about their own relationships with better information while the decisions remain theirs.

## References

- Asendorpf, J. B., Penke, L., & Back, M. D. (2011). From dating to mating and relating: Predictors of initial and long-term outcomes of speed-dating in a community sample. *European Journal of Personality*, 25(1), 16–30. <https://doi.org/10.1002/per.768>
- Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, 66(3), 411–421. <https://doi.org/10.1097/01.psy.0000127692.23278.a9>
- Back, M. D., Schmukle, S. C., & Egloff, B. (2011). A closer look at first sight: Social relations lens model analysis of personality and interpersonal attraction at zero acquaintance. *European Journal of Personality*, 25(3), 225–238. <https://doi.org/10.1002/per.790>
- Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, 8(3), 338–363. <https://doi.org/10.1037/1082-989X.8.3.338>
- Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, 106(2), 265–289. <https://doi.org/10.1037/0033-2909.106.2.265>
- Bruch, E. E., & Newman, M. E. J. (2018). Aspirational pursuit of mates in online dating markets. *Science Advances*, 4(8), eaap9815. <https://doi.org/10.1126/sciadv.aap9815>
- Butler, E. A. (2011). Temporal interpersonal emotion systems: The "TIES" that form relationships. *Personality and Social Psychology Review*, 15(4), 367–393. <https://doi.org/10.1177/1088868311411164>
- Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, 5(2), 153–163. <https://doi.org/10.1089/big.2016.0047>
- Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD statement. *Annals of Internal Medicine*, 162(1), 55–63. <https://doi.org/10.7326/M14-0697>
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187–202. <https://doi.org/10.1111/j.2517-6161.1972.tb00899.x>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34(7), 571–582. <https://doi.org/10.1037/0003-066X.34.7.571>
- Eastwick, P. W., & Hunt, L. L. (2014). Relational mate value: Consensus and uniqueness in romantic evaluations. *Journal of Personality and Social Psychology*, 106(5), 728–751. <https://doi.org/10.1037/a0035884>
- Eastwick, P. W., Finkel, E. J., Mochon, D., & Ariely, D. (2007). Selective versus unselective romantic desire: Not all reciprocity is created equal. *Psychological Science*, 18(4), 317–319. <https://doi.org/10.1111/j.1467-9280.2007.01897.x>
- Eastwick, P. W., Luchies, L. B., Finkel, E. J., & Hunt, L. L. (2014). The predictive validity of ideal partner preferences: A review and meta-analysis. *Psychological Bulletin*, 140(3), 623–665. <https://doi.org/10.1037/a0032432>
- Felmlee, D. H., & Greenberg, D. F. (1999). A dynamic systems model of dyadic interaction. *The Journal of Mathematical Sociology*, 23(3), 155–180. <https://doi.org/10.1080/0022250x.1999.9990218>
- Finkel, E. J., Eastwick, P. W., Karney, B. R., Reis, H. T., & Sprecher, S. (2012). Online dating: A critical analysis from the perspective of psychological science. *Psychological Science in the Public Interest*, 13(1), 3–66. <https://doi.org/10.1177/1529100612436522>

- Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, 3(4), 456–465. <https://doi.org/10.1177/2515245920952393>
- Fraley, R. C. (2002). Attachment stability from infancy to adulthood: Meta-analysis and dynamic modeling of developmental mechanisms. *Personality and Social Psychology Review*, 6(2), 123–151. [https://doi.org/10.1207/S15327957PSPR0602\\_03](https://doi.org/10.1207/S15327957PSPR0602_03)
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/b16018>
- Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378. <https://doi.org/10.1198/016214506000001437>
- Gottman, J. M., & Levenson, R. W. (1992). Marital processes predictive of later dissolution: Behavior, physiology, and health. *Journal of Personality and Social Psychology*, 63(2), 221–233. <https://doi.org/10.1037/0022-3514.63.2.221>
- Gottman, J. M., & Levenson, R. W. (2000). The timing of divorce: Predicting when a couple will divorce over a 14-year period. *Journal of Marriage and Family*, 62(3), 737–745. <https://doi.org/10.1111/j.1741-3737.2000.00737.x>
- Gottman, J. M., Coan, J., Carrere, S., & Swanson, C. (1998). Predicting marital happiness and stability from newlywed interactions. *Journal of Marriage and the Family*, 60(1). <https://doi.org/10.2307/353438>
- Gottman, J., Swanson, C., & Murray, J. (1999). The mathematics of marital conflict: Dynamic mathematical nonlinear modeling of newlywed marital interaction. *Journal of Family Psychology*, 13(1), 3–19. <https://doi.org/10.1037/0893-3200.13.1.3>
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12(1), 19–30. <https://doi.org/10.1037/1040-3590.12.1.19>
- Hazan, C., & Shaver, P. (1987). Romantic love conceptualized as an attachment process. *Journal of Personality and Social Psychology*, 52(3), 511–524. <https://doi.org/10.1037/0022-3514.52.3.511>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Heyman, R. E., & Slep, A. M. S. (2001). The hazards of predicting divorce without crossvalidation. *Journal of Marriage and Family*, 63(2), 473–479. <https://doi.org/10.1111/j.1741-3737.2001.00473.x>
- Joel, S., Eastwick, P. W., & Finkel, E. J. (2017). Is romantic desire predictable? Machine learning applied to initial romantic attraction. *Psychological Science*, 28(10), 1478–1489. <https://doi.org/10.1177/0956797617714580>
- Joel, S., Eastwick, P. W., Allison, C. J., Arriaga, X. B., Baker, Z. G., Bar-Kalifa, E., Bergeron, S., Birnbaum, G. E., Brock, R. L., Brumbaugh, C. C., Carmichael, C. L., Chen, S., Clarke, J., Cobb, R. J., Coolsen, M. K., Davis, J., de Jong, D. C., Debrot, A., DeHaas, E. C., . . . Wolf, S. (2020). Machine learning uncovers the most robust self-report predictors of relationship quality across 43 longitudinal couples studies. *Proceedings of the National Academy of Sciences*, 117(32), 19061–19071. <https://doi.org/10.1073/pnas.1917036117>
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4), 237–251. <https://doi.org/10.1037/h0034747>
- Karney, B. R., & Bradbury, T. N. (1995). The longitudinal course of marital quality and stability: A review of theory, method, and research. *Psychological Bulletin*, 118(1), 3–34. <https://doi.org/10.1037/0033-2909.118.1.3>
- Kenny, D. A., & La Voie, L. (1984). The social relations model. *Advances in Experimental Social Psychology*, 18, 141–182. [https://doi.org/10.1016/S0065-2601\(08\)60144-6](https://doi.org/10.1016/S0065-2601(08)60144-6)
- Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. Guilford Press.

- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15), 5802–5805. <https://doi.org/10.1073/pnas.1218772110>
- Kurdek, L. A. (2004). Are gay and lesbian cohabiting couples really different from heterosexual married couples? *Journal of Marriage and Family*, 66(4), 880–900. <https://doi.org/10.1111/j.0022-2445.2004.00060.x>
- Lakens, D., & Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science*, 4(1), 2515245920951503. <https://doi.org/10.1177/2515245920951503>
- Le, B., Dove, N. L., Agnew, C. R., Korn, M. S., & Mutso, A. A. (2010). Predicting nonmarital romantic relationship dissolution: A meta-analytic synthesis. *Personal Relationships*, 17(3), 377–390. <https://doi.org/10.1111/j.1475-6811.2010.01285.x>
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. University of Minnesota Press. <https://doi.org/10.1037/11281-000>
- Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*, 4(2), 139–157. <https://doi.org/10.1037/1082-989X.4.2.139>
- Norris, J. R. (1997). *Markov chains*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511810633>
- Norton, M. I., Frost, J. H., & Ariely, D. (2007). Less is more: The lure of ambiguity, or why familiarity breeds contempt. *Journal of Personality and Social Psychology*, 92(1), 97–105. <https://doi.org/10.1037/0022-3514.92.1.97>
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606. <https://doi.org/10.1073/pnas.1708274114>
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., & Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15), 1923–1929. <https://doi.org/10.1093/bioinformatics/btp358>
- Reis, H. T., Maniaci, M. R., Caprariello, P. A., Eastwick, P. W., & Finkel, E. J. (2011). Familiarity does indeed promote attraction in live interaction. *Journal of Personality and Social Psychology*, 101(3), 557–570. <https://doi.org/10.1037/a0022885>
- Salganik, M. J., Lundberg, I., Kindel, A. T., Ahearn, C. E., Al-Ghoneim, K., Almaatouq, A., Altschul, D. M., Brand, J. E., Carnegie, N. B., Compton, R. J., Datta, D., Davidson, T., Filippova, A., Gilroy, C., Goode, B. J., Jahani, E., Kashyap, R., Kirchner, A., McKay, S., . . . McLanahan, S. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences*, 117(15), 8398–8403. <https://doi.org/10.1073/pnas.1915006117>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Singer, J. D., & Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics*, 18(2), 155–195. <https://doi.org/10.3102/10769986018002155>
- Stanley, S. M., Bradbury, T. N., & Markman, H. J. (2000). Structural flaws in the bridge from basic research on marriage to interventions for couples. *Journal of Marriage and Family*, 62(1), 256–264. <https://doi.org/10.1111/j.1741-3737.2000.00256.x>

- Strathern, M. (1997). 'Improving ratings': Audit in the British University system. *European Review*, 5(3), 305–321. [https://doi.org/10.1002/\(SICI\)1234-981X\(199707\)5:3<305::AID-EURO184>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1234-981X(199707)5:3<305::AID-EURO184>3.0.CO;2-4)
- Strogatz, S. H. (1988). Love affairs and differential equations. *Mathematics Magazine*, 61(1), 35. <https://doi.org/10.1080/0025570X.1988.11977342>
- Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., & Steyerberg, E. W. (2019). Calibration: The Achilles heel of predictive analytics. *BMC Medicine*, 17(1), 230. <https://doi.org/10.1186/s12916-019-1466-7>
- Visser, I. (2011). Seven things to remember about hidden Markov models: A tutorial on Markovian models for time series. *Journal of Mathematical Psychology*, 55(6), 403–415. <https://doi.org/10.1016/j.jmp.2011.08.002>
- Whitchurch, E. R., Wilson, T. D., & Gilbert, D. T. (2011). "He loves me, he loves me not . . .": Uncertainty can increase romantic attraction. *Psychological Science*, 22(2), 172–175. <https://doi.org/10.1177/0956797610393745>
- Woodlock, D. (2017). The abuse of technology in domestic violence and stalking. *Violence Against Women*, 23(5), 584–602. <https://doi.org/10.1177/1077801216646277>
- Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12(6), 1100–1122. <https://doi.org/10.1177/1745691617693393>

---

## NOTICE

This whitepaper is published by the Institute for Human Evolution for educational and research purposes. It summarises and evaluates published scientific literature and proposes directions for future research. It is not medical, psychological, legal, or relationship advice, it does not describe services that IHE provides to individuals, and it should not be used to diagnose or treat any condition. IHE's work is preventive education and skill-building, not treatment.

If you are struggling after a loss or a breakup, a licensed mental-health professional can help. In the United States, the 988 Suicide & Crisis Lifeline is available by calling or texting 988. If you are in immediate danger, contact local emergency services.

This paper has not undergone external peer review. Citations were checked against published records at the time of writing; corrections are welcome at [hello@ihe.institute](mailto:hello@ihe.institute).

© 2026 Institute for Human Evolution. 5449 Hollywood Blvd Unit A, Los Angeles, CA 90027. [hello@ihe.institute](mailto:hello@ihe.institute)