Capstone Project Proposal

Matchings and Dating App Market

Group 4: Corneel Henri L. Moons, Jillian Ann Hunter, Sofia Pirogova

1 Introduction

In the age of technology, digital dating has shifted from a niche option to a dominant pathway for forming new relationships. With potential matches only a swipe away, users are immersed in larger and more competitive dating pools than ever before, raising questions about how to present themselves and whom to pursue to achieve their desired outcome. Prior to online platforms, people adapted traits and behaviors to relate to those they were attracted to (Toledano et al., 2025), but even then, successful interaction depended on access to information. Today, that information, such as who has already "liked" you, is selectively revealed: some apps provide it freely, while others hide it behind a paywall.

Even on platforms that collect extensive profile data, individuals remain uncertain about how many potential partners share their relationship goals, how competitive their local dating market is, or how likely higher-rated users are to reciprocate. This information design shapes beliefs about one's own desirability, expected acceptance probabilities, and optimal swiping and self-presentation strategies. Understanding how these information structures influence behavior is central to studying platform design, matching efficiency, and consumers' willingness to pay for premium visibility features.

In this section we formalize a simple decision problem that links users' beliefs, effort choices, and the observable OkCupid profiles in our cross-sectional data.

2 Conceptual Framework

We study how users of a dating app update their beliefs about the market from the profiles they see, become frustrated, and then adjust their behaviour (effort, presentation, or exit). The model is deliberately simple and geared towards simulation and comparative statics.

2.1 Agents and Market

Each user i is characterised by:

- a *relationship goal* $g_i \in \{S, L\}$, where S denotes short-term and L long-term orientation;
- a *market rating* $r_i \in [0, 1]$ summarising how desirable the user is as a partner (looks, education, lifestyle, etc.).

Users interact in markets m defined by:

- a true share ρ_m of long-term–oriented users;
- a distribution $F_m(r \mid g)$ of ratings by type $g \in \{S, L\}$.

Users have aspiration levels: a user with rating r_i prefers partners with rating $r \geq r_i$ and whose goals are compatible with g_i. High-rating long-term users are especially selective; short-term users are willing to match with a wider range of partners.

In the empirics, r_i will be proxied by a scalar index based on observable characteristics (demographics, education, lifestyle, profile completeness), and ρ_m by the observed share of profiles stating long-term intent in market m.

2.2 Bayesian Learning

A newcomer arrives with imperfect information about the market and her own prospects. She holds prior beliefs about:

- the share of serious users in the market, ρ_m;
- her success probability with *acceptable* partners (goal-compatible and sufficiently high rating), denoted p_i.

Upon joining, the app shows her a finite sample of K profiles drawn from the current market composition. From this sample she observes:

- the share of profiles that look long-term–oriented (stated intent, cues in the profile);
- the distribution of ratings among these profiles;

Using these signals, the user updates her beliefs to posteriors $\widehat{\rho}_m$ and $\widehat{p}_i$ via Bayesian learning: seeing many casual-looking or low-effort profiles and receiving few positive responses induces a lower $\widehat{\rho}_m$ and a lower $\widehat{p}_i$

$$b_i^* \in \arg\max_{b_i \in \{S,A\}} \left\{ \mathbb{E}\big[u_i(r_j, g_j) \,\big|\, r_j \geq r_i,\ g_j \sim g_i,\ \widehat{\rho}_m, F_m, b_i\big] - C_i(b_i) \right\}$$

. We interpret this joint pessimism as *frustration*: the belief that the market is dominated by the "wrong" types and that success with truly desirable partners is unlikely.

This generates a dynamic feedback loop:

1. High-rating long-term users become frustrated and reduce effort or leave the app.
2. The visible pool of profiles becomes more dominated by short-term or ambiguous profiles and by low-effort or "too good to be true" signals.

3. Newcomers see this pool, update to more pessimistic $(\widehat{\rho}_m, \widehat{p}_i)$, and themselves are more likely to exit.

We use simulations to trace how the composition of the market evolves under different initial conditions and cost–benefit parameters, with particular attention to the share of high-rating long-term users who remain engaged.

3 Data Plan

3.1 Dataset

We use the OkCupid dataset containing approximately 60,000 profiles, including structured demographics, multiple categorical and continuous attributes, and up to ten essay fields. We treat each profile as a cross-sectional snapshot of the user's effort choice E_i and representation choice H_i after some (unobserved) time on the app.

Concept	**Proxy from Data**
Effort E_i	Essay word count; number of essays completed; profile completeness
Representation H_i	Stated goals (LTR vs casual vs ambiguous); patterns in "idealised" traits
Market seriousness	Share of profiles in same location × orientation indicating long-term intent
Scarcity / market tightness	Opposite-sex to same-sex ratio per orientation group
Signal clarity	Average completeness of partner profiles in the user's group

Representation H_i will be proxied both by how clearly the user states long-term vs casual goals and by the presence of attributes that are likely to be embellished (e.g. heaping at "round" heights, unusually high shares of "never" drinking in groups where this is atypical).

3.2 Market-Level Indices

For each location × orientation group we compute:

- the share of profiles signalling long-term intent (proxy for market seriousness $\widehat{\rho}_m$),
- the average profile completeness among potential partners (signal clarity),
- the male/female ratio (or analogous competition metric, proxy for scarcity).

These indices are merged back to individual-level observations and used as proxies for the belief environment described in the conceptual framework.

3.3 Testable Predictions

We use the model for two comparative-static exercises:

1. **Where are we now?** Given the estimated fundamentals of a market, we compare the observed share of high-effort long-term profiles to the share predicted by a frictionless benchmark. A deficit is interpreted as evidence of frustration and inefficient use of the platform.

2. **What can the platform do?** We study simple design policies, such as which types of profiles are shown in the *first* sample K that newcomers see. Because beliefs evolve via Bayesian learning, curating early exposure can shift $\widehat{\rho}_m$ and $\widehat{p}_i$, keeping high-rating long-term users in Strategy S for longer and reducing premature exit.

The following predictions follow directly from the Bayesian belief-updating framework with frustration and subsequent adaptation of effort and representation choices:

1. **Market pessimism effect:** Users in markets with a low share of long-term–intent profiles and high competition exert lower effort and are more likely to present ambiguous or casual goals.

2. **Adaptation / embellishment effect:** In such markets, the probability of adapted presentation (lless sincere, more embellished profiles) is higher, reflected in a greater prevalence of "too good to be true" combinations of attributes.

3. **Information clarity effect:** Markets with more complete partner profiles (higher signal clarity) exhibit lower dispersion in effort conditional on user characteristics, consistent with clearer beliefs about the market.

4. **Interaction effect:** Scarcity amplifies pessimism; crowded, casual markets (low seriousness, high competition) yield the lowest effort levels and the highest incidence of adapted / embellished profiles.

References

[1] Toledano, N., Viejo, C., & Ortega-Ruiz, R. (2025). *Romantic Competence and Courtship Skills: From the First Romantic Impulse to the Management of Mutuality. Journal of Adolescence*, 97(4), 1047–1056.