

# “Hacking” OkCupid

## One Man’s Extremely Work-Intensive Path to Love

Clara Bennett

April 30, 2014

# What's an OkCupid, Anyway?

- Free online dating service
- Founded in 2004 by a team of four Harvard math grads: Chris Coyne, Sam Yagan, Max Krohn, and Christian Rudder
- Utilizes a matching algorithm that leaves the domain knowledge to the users
- Popular amongst Millennials, hipsters, nerds<sup>1</sup>

---

<sup>1</sup>This claim is based on artisanal data  
2 / 11

# How OkCupid Does Matching

- User-generated, multiple-choice match questions
  - Response
  - Desired response(s) for potential matches
  - Importance rating (weighting: 0, 1, 10, 50, 250)
- An algorithm determines the “goodness” of match questions, based on average importance rating and diversity of the responses
- Users are presented with the best questions first, and may answer as many questions as they wish

## How OkCupid Does Matching, Now with Math!

- From the set,  $S$ , of questions answered by Alice and Bob:
  - Calculate the percent,  $P_{A \rightarrow B}$ , of Alice's responses that are acceptable to Bob, weighted by importance, and vice versa

$$\begin{aligned}M_{raw} &= \sqrt{P_{A \rightarrow B} P_{B \rightarrow A}} \\ M &= M_{raw} - \frac{1}{|S|}\end{aligned}$$

- TL;DR: Importance rating makes a big difference, and having lots of questions answered in common increases the highest potential score.

## The Man Himself



*Photo: Emily Shur, from Wired*

- Chris McKinlay, PhD in mathematics from UCLA, former card-counting blackjack player
- In 2012, McKinlay was working on his dissertation (large-scale data processing and parallel numerical methods) and struggling to find a special someone online
- Few match questions in common with women in his area → low match percentages → low visibility

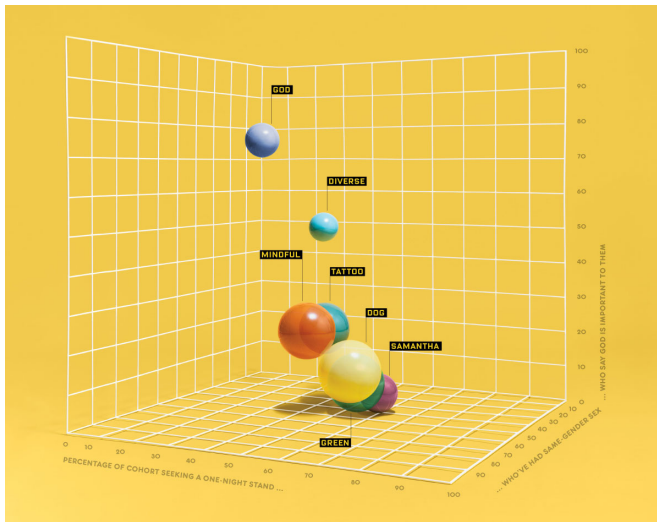
## Date Like a Mathematician

- McKinlay wanted to boost his compatibility ratings with the women whom he was interested in meeting
- To do so, McKinlay decided to play to his strengths by:
  - Collecting match data on his target demographic, via 12 bot profiles
  - Extracting the match questions that were most important to the women in whom he'd most likely be interested
  - Crafting a new profile that answered all the right match questions

# Collection and Analysis

- The first pass, nation-wide:
  - 3 weeks, using bots programmed to mimic human activity
  - 20,000 women
  - 6,000,000 match question answers
- Analyzed using K-modes clustering, settling on a solution with 7 distinct clusters
- The second pass, Los Angeles and San Francisco:
  - 5000 women
  - Confirmed clustering pattern from first sample
- By this point, apparently, McKinlay was sleeping on his desk

# Clusters Full of Women





## Applying the Results

- McKinlay browsed the clusters and chose two to target, optimizing a profile for each
- Text-mined the profiles to determine common interests, so he'd know which traits to emphasize
- Answered (honestly) the 500 most popular match questions for each cluster, using adaptive boosting techniques to determine how important to rate them

## Spoiler Alert: He Finds Love in the End

- As a result of his efforts, McKinlay ended up matching better than 90% with over 10,000 women in the LA area
- Wrote a program to auto-visit his top matches, resulting in tons of incoming messages from interested women
- Went on 88 first dates, sometimes multiple in one day, before finding the woman to whom he would become engaged

## References and Links

- Mathematic description of matching algorithm:  
<https://www.okcupid.com/help/match-percentages>
- Original article in Wired:  
<http://www.wired.com/2014/01/how-to-hack-okcupid/all/>
- OkCupid data blog (discontinued, but old posts are interesting):  
<http://blog.okcupid.com/>