ROMANTIC REGRESSIONS

An Analysis of Behavior in Online Dating Systems

Andrew Rocco Tresolini Fiore
A.B., College Scholar
Cornell University
2002

Submitted to the
Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of
Master of Science in Media Arts and Sciences
at the Massachusetts Institute of Technology

September 2004

Author
Program in Media Arts and Sciences
September 2004

Certified by
Judith S. Donath
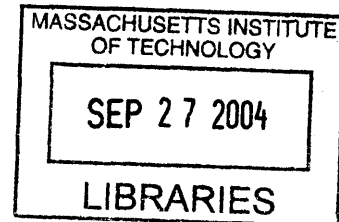Assistant Professor of Media Arts and Sciences
Thesis Supervisor

Accepted by
Andrew B. Lippman
Chair, Departmental Committee on Graduate Studies
Program in Media Arts and Sciences

ROMANTIC REGRESSIONS

An Analysis of Behavior in Online Dating Systems
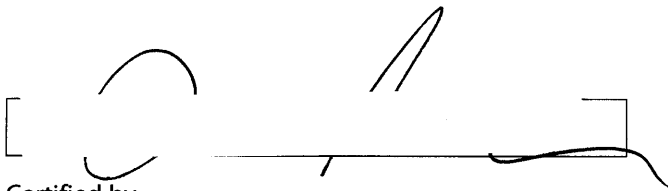
by
Andrew Rocco Tresolini Fiore

ABSTRACT

Online personal advertisements have shed their stigma as matchmakers for the awkward to claim a prominent role in the social lives of millions of people. Web sites for online dating allow users to post lengthy personal ads, including text and photos; search the database of users for potential romantic partners; and contact other users through a private messaging system.

This work begins with psychological and sociological perspectives on online dating and discusses the various types of online dating Web sites. Next, it presents an analysis of user behavior on one site in particular, which has more than 57,000 active users from the United States and Canada. A demographic description of the population is given, and then 250,000 messages exchanged by the active users over an eight-month period are analyzed. An examination of which characteristics are "bounding" finds that life course attributes such as marital status and whether one wants children are most likely to be the same across the two users in a dyadic interaction. To understand which characteristics are important to users in deciding whom to contact, regression models show the relative strength of a variety of attributes in predicting how many messages a user with those attributes will receive. By far the strongest predictor of *messages received* is the number of messages sent. For men, age, educational level, and self-rated physical attractiveness are the next most important qualities. For women, they are not being overweight, self-rated physical attractiveness, and having a photo. Finally, a discussion of the design implications of these findings and other design issues follow the results.

Thesis Supervisor: Judith S. Donath, Assistant Professor of Media Arts and Sciences

Certified by

Dan Ariely

Luis Alvarez Renta Professor of Management Science

Thesis Reader

Certified by

Pattie Maes

Associate Professor of Media Arts and Sciences

Thesis Reader

ACKNOWLEDGEMENTS

Chapter 1

Introduction to Online Dating

Human and institutional intermediaries for centuries have connected couples for romantic relationships. Match-makers, video dating, newspaper personal advertisements, and speed dating vie for the faith and the money of their clients (Ahuvia & Adelman 1992).

These services have historically served only a small subset of the single population; the stigma of desperation and social awkwardness pervades popular accounts of their users (Ahuvia & Adelman 1992). Perhaps the stigma was not entirely incorrect: Goodwin (1990) found that although users of a dating service in Britain equalled a control group in socioeconomic status, they lagged behind in dating skills. It makes sense that the small, elective population of a dating service would be self-selected for having difficulty meeting a romantic partner through traditional channels.

ONLINE PERSONAL ADVERTISEMENTS

Online personal advertisements — lengthier, more detailed descendants of newspaper personal ads — have grown rapidly in recent years. In August 2003, personals Web sites in the United States drew 40 million unique visitors — half the number of single adults in the U.S. (Mulrine 2003). (Chances are that not all of the visitors were single, however.) With that many users, attributes tend to regress to the mean across the population: it is clear that the base of users no longer comprises only the lonely, socially stilted group that, according to the stereotype, populated traditional personal ads.

Thus, online personal ads have shed the stereotype as matchmakers for the awkward and now claim a prominent role in the social lives of millions of users. Millions more have friends and relatives who have used these systems, helping to personalize and destigmatize the public image of someone who would place a personal ad. From popular accounts, the base of users spans generations, breaking the 20- or 30-something age ceiling common in many online social environments.

Researchers have studied online friendships and romantic relationships from psychological and sociological perspectives (Van Gelder 1985, Lea & Spears 1995, Walther 1996, McKenna et al. 2002), and psychologists and sociologists have studied the personals ads that appear in print publications (Bolig et al. 1984, Ahuvia & Adelman 1992). To date, however, online personals have been considered only anecdotally. Their careful study is vital for two reasons.

First, as a suddenly popular way to meet partners for dates or relationships, online personals not only reflect but also have the potential to shape how people attract one another, date, and fall in love. The design of social systems influences the beliefs and behavior of their users; the features of a person that Match.com presents as salient to romance will begin to have some psychological and cultural influence if 40 million Americans view them every month. This influence exerts itself most strongly on the users themselves, but it also spreads into the general population. Because online personals systems affect such an important and intimate arena of their users' lives, designers have a particular obligation to work carefully.

Second, usage data from online personals Web sites provides an opportunity to perform rich analyses of the preferences and behavior of people seeking relationships. This information is intrinsically interesting as social science, but, more importantly, it will also inform future designs and matching algorithms to improve the experience of online dating.

AMERICAN SOCIETY AND THE RISE OF ONLINE DATING

The increasing ubiquity of Internet access in the United States has allowed a large number of users to reach online dating systems; the presence of this critical mass makes the systems increasingly attractive to even more users. However, a variety of other factors may have facilitated the sudden rise to prominence of online dating as a way to meet partners.

Attitudes Toward Marriage

The notion of a love marriage — in which a partner is selected to maximize emotional satisfaction rather than for socioeconomic or familial reasons — is an historically recent idea (Fisher 1992). It is also largely a Western one. Because romantic compatibility is difficult to predict and more fickle than practical compatibility (e.g., shared interests, socioeconomic status), those seeking a love marriage may be more selective than others and more willing to seek outside help (Ahuvia & Adelman 1992). The ability provided by online dating to sift quickly through tens of thousands of potential mates caters well to the desire of users to find a potentially elusive love match.

The ability to search almost endlessly may also support the cultural notion of the "one true love" — the idea that only one person will make you truly happy.

Delayed Marriage and Increased Mobility

Marriage tends to take place later in the life course than it did in the past; in just two decades, from 1975 to 1996, the median age of first marriage rose from 20.4 to 21.3 years for women and 22.9 to 23.5 years for men (U.S. Census Bureau 1975, 2002). Because more people are marrying after they leave natural dating pools like high school and college, they must employ other methods to meet potential partners. Finding partners outside of these institutional venues is a more difficult task, so formal intermediaries like online personals become a more valuable resource.

Furthermore, the American prioritization of career leads many people to move to distant cities for work or education, to places where they have no friends and no family. People who make moves like this must rebuild local social ties from nothing, a relatively rare situation for most people in most parts of the world.

Online personal ads, for finding dates or just friends, provide a mechanism for users to establish new connections even without an existing basis from which to expand their social circle.

Divorce Rate

The divorce rate in the U.S. is among the highest in the world, at 4.5 per 1,000 population; in most nations the rate is under 3.0 per 1,000 (United Nations 1999). Whatever its causes, the relatively high divorce rate in the U.S. creates a relatively large population of single people in mid-life, sometimes with children at home — certainly young enough to seek another life partner, but often too busy with children or career to spend much time broadening their social circles to find other single people. This population might represent a significant portion of

the boom in people using online dating systems who are older than typical computer-mediated communication users.

A Reciprocal Relationship with Culture

Social technologies are rarely neutral with respect to the interactions that take place through them. By their form and function, they affect the way their users communicate. Gay and Hembrooke (2004) propose an ecological extension of activity theory to "describe and account for the transformative power of seemingly ubiquitous artifacts such as language and pervasive computing devices":

> "Mutual accommodations among system elements shape the relationship among these components, which is interdependent. Changes in any part of a system or among contextual levels have the potential to affect any or all of the other related systems. ... Activities such as technology construction should not be perceived as statically structured entities but rather as dynamic processes that are characterized by ambiguity and change."

> (Gay and Hembrooke 2004)

The designer of a personals system chooses to emphasize (both visually and algorithmically) some characteristics more than others. These decisions affect users' perceptions of what is important in others they see on the site. More broadly, I believe the choice of characteristics of people to emphasize in an online dating system has a reciprocal relationship with cultural notions of love and attraction (Figure 1.1).

As a suddenly popular way to meet partners for dates or relationships, online personals not only reflect but also have the potential to shape how people attract one another, date, and fall in love. The design of social systems influences the beliefs and behavior of their users; the features of a person that Match.com presents as salient to romance will begin to have some psychological and cultural influence if tens of millions of Americans view them every month. This influence would exert itself most strongly on the users themselves, but it could also spread into the general population.

Users will perceive a system as more efficacious if its design highlights the characteristics of others that are truly
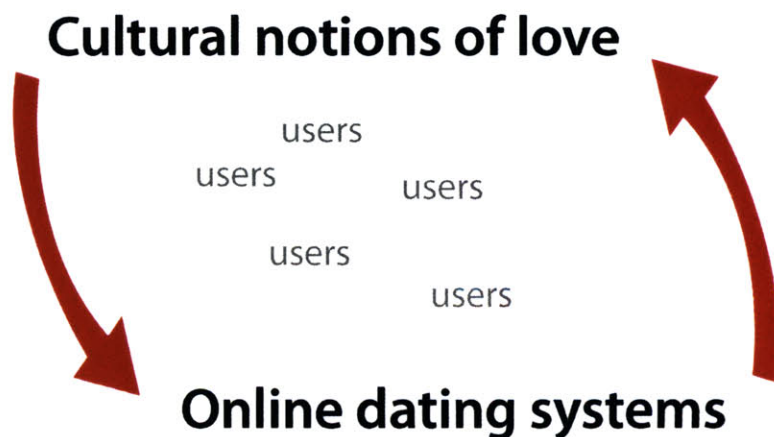
Cultural notions of love

users
users   users
users
users

Figure 1.1. The reciprocal relationship hypothesized between culture and online dating systems.

Online dating systems

15

HOW ONLINE PERSONALS WORK

Online personals systems usually include personal profiles of users, which document their location, gender, age, physical attributes, race, religion, smoking and drinking habits, self-description, and preferences for these characteristics in a potential mate or date; a private messaging system; and a mechanism for indicating interest without writing a substantial message, variously termed "winking," "eye contact," or a "collect call." In contrast to other systems that support online communities, personals systems usually lack a common forum, where everyone can read what anyone posts.

Profiles

User profiles include a title or pseudonym, constrained descriptors like age and hair color, free-response descriptors, and sometimes photographs. Constrained descriptors limit users to certain categories (e.g., the set of hair colors) or types of input (e.g., numeric for age). Free-response descriptors let the user answer questions with arbitrary text. Some sites give straightforward prompts for free-response descriptors, such as "Describe yourself" or "Describe who you're looking for." Others, such as Spring Street Networks (provider for Nerve.com and other sites), provide less direct prompts: "Best (or worst) lie I've ever told," "Song or album that puts me in the mood."

On Nerve.com, frequent users have become so familiar with common responses to the idiosyncratic free-response prompts that they sometimes make mocking reference to clichéd answers to a question in their own response.

Some new personals sites incorporate personality psychology into their profile-building process; Tickle, eHarmony, and PerfectMatch purport to match their members according to the results of personality tests.

Searching and Matching

Most systems provide both searching and matching functionality. In *searching*, users specify exactly what characteristics they are looking for. They can search the set of profiles based on constrained descriptors such as age, eye color, and religion, and sometimes by keywords in the free-response descriptors.

In *matching*, which is slightly less direct, systems pair users by comparing their profile descriptors to the descriptors of others — usually the constrained descriptors, because contemporary techniques for clustering or otherwise identifying similarity work better with clearly defined features than with free text. Different systems apply different weightings to the features of the profiles.

Private Messaging

Online personals systems invariably provide a private channel through which members communicate without revealing their names or regular email addresses. Typically, it functions like a Web-based email client, providing a field to enter a subject for the message and a larger field for the message body, except that users can write only to others on the same site.

Users generally write each other with the private messaging system to find out whether they want to proceed to contact via email, phone, or face-to-face interaction. This provides privacy and safety until the users are comfortable with each other. A smaller group of users maintains contact for an extended period via the private messaging system, either uninterested in migrating the relationship to another medium or unwilling to do so.

Although sites often allow free browsing and searching in order to attract customers, most require users to pay for the ability to send private messages. Some sites allow non-paying users to reply to messages from paying users but not to initiate conversations on their own.

eHarmony restricts whom even paying customers can contact. In this system, users can contact a potential date only after the system has matched them with the person based on the results of their personality tests. That is, eHarmony provides only matching, not searching, and prevents communication between people who are not matched by the system.

Additionally, some personals systems offer a form of free contact without a substantial message, variously called an "eye contact," a "wink," or a "collect call." These mechanisms allow a user to express interest in another without formulating a message or, usually, paying for membership. On Nerve, some women are so deluged with collect calls that they indicate in their profiles that they will not respond to them, often criticizing the stinginess of those who do not pay for full privileges but still want a response.

THE CONSTRUCTION OF IDENTITY

Because text-based media provide only a limited communicative modality, it is possible to control what one conveys in such media more fully than in face-to-face interaction. In the language of Goffman (1959), users can tailor their self-presentation so completely that they accidentally "give off" nothing, communicating only what they intend to "give." This level of control allows users to construct and maintain one or more personae easily and convincingly:

> "In the physical world there is an inherent unity to the self, for the body provides a compelling and convenient definition of identity. The norm is: one body, one identity. Though the self may be complex and mutable over time and circumstance, the body provides a stabilizing anchor. [...] The virtual world is different. It is composed of information rather than matter. [...] One can have, some claim, as many electronic personas as one has time and energy to create."

> (Donath 1998)

Personal ads are clearly a venue in which favorable self-presentation is crucial. This makes the ability to craft one's identity particularly salient; many users expend great effort in perfecting their profiles, tweaking small bits of wording to achieve just the right effect. Because the medium is asynchronous, users can spend hours creating a profile that will be read in only a minute or two.

The Strategy of Self-Presentation

Online dating systems function as economies of attention. Users compete for the time and effort of others to read

their profiles and perhaps respond; this promotes what we might call strategic communication, in which people tailor their self-presentation to display what they believe will help them achieve their objectives. One could claim that all communication is strategic to some degree, and in dating situations most people want to put their best face forward. However, in online dating, the potential for exaggeration or fabrication is much greater. At 5'8", I cannot successfully claim to a person in my physical presence that I am 6'2" tall. But in my online personal ad, the deceit would be hard to detect.

Text-based online personals facilitate exaggeration. They make it easy to embellish or lie about characteristics that would be hard to misrepresent in face-to-face interaction, such as height, weight, age, attractiveness, and even facial appearance. (Even photographs are not unimpeachable: they might be favorable shots of a person who usually appears less attractive, or they might show someone else entirely.)

Once a significant number of users in a given personals system post profiles with exaggerated personal information, other users have an incentive to exaggerate to a similar degree or risk being overlooked in favor of others with more appealing — but not necessarily true — attributes (Ariely 2003). This incentive promotes strategic self-presentation, in which users attempt to present a profile, even enact a persona, which they think will optimize their desirability to others, regardless of whether it represents them accurately.

Strategic self-presentation is easier in text-based media than in face-to-face interaction because the user can strictly control what information she gives — that which she intentionally conveys — and can nearly eliminate what she "gives off" — that which she inadvertently indicates (Goffman 1959). Few people can avoid "giving off" their sex or height in face-to-face interaction.

Even if we accept that dating situations often involve some exaggeration in the service of self-promotion, the degree to which it is possible through informationally impoverished channels, combined with the incentive to exaggerate in an environment of exaggerators, makes online dating users likely to exaggerate more often and to a greater degree than offline daters.

Perception of Others

With such care devoted to production of profiles, their consumption requires interpretation beyond literal comprehension. Users of online dating systems must learn through experience — or perhaps from the advice of those more experienced — how to read between the lines in the local culture, which varies from site to site.

The Culture of a Personals Site

On Nerve.com, which shares content with other affiliates of Spring Street Networks' personals system, users self-consciously style hip identities. Certain bands, artists, and writers become points of reference, examples of various identities. Nerve's free-text profile questions seem designed to provoke this kind of expression: "Best (or worst) lie I've ever told," "Song or album that puts me in the mood." The prompt, "Most humbling moment," provokes a common response, "Posting this ad," but also commentary on it. (A paraphrased example: If posting this ad is your most humbling moment, you need to get out more.) This kind of explicit meta-commentary is rare, but users

commonly demonstrate familiarity with the norms, if only through their invocation of the familiar examples.

Vaisman (2001) offers a send-up of the stereotypical Nerve user:

> "You have a dog-eared copy of Stendhal, in French, next to the iMac and the Dutch lube you picked up in Amsterdam. You dream of your next trip to Iceland. You try your hand at pathetic fallacies: *your fingers itch, the trees shudder*. Belle & Sebastian are pretty awesome. You work for a website, but you know *what's up*."

> (Vaisman 2001)

The identities of the various online dating sites arise from a combination of their marketing with the real attributes of their populations. eHarmony.com, which claims to match users based on personality compatibility, focuses more on marriage than most sites; it sometimes bills itself as a Christian match-making service. Nerve Personals (and Spring Street Networks) arose from an eponymous online magazine about sex. As such, its marketing evokes dating and sex more than marriage. Some sites, like Match.com and Yahoo! Personals, seem less targeted, with marketing that presents nonspecific images of romance.

Social Affinity, Real and Imagined

Walther et al. postulate that the tendency in computer-mediated communication to fill in the blanks about a conversational partner more optimistically than one would in face-to-face interaction give the media a "hyperpersonal" effect, in which participants project desirable qualities onto their conversational partners in lieu of complete information about them (Walther 1996; Walther et al. 2001). The feeling of social affinity is quickened, but this sense comes more from wishful thinking than from accurate perception of the other, so it can lead to disappointment later, when a fuller picture of the other has emerged.

Walther's early development of hyperpersonal theory (1996) found that text-based media can facilitate social affinity, but the affinity takes longer to bloom than it would via face-to-face interaction, because conversation is less rapid and extraverbal cues are largely absent. Thus, computer-mediated interaction allows the formation of intimacy but slows the process as compared to face-to-face communication because the conveyance of salient information is slower.

His later work with collaborators (Walther et al. 2001) delineated a more nuanced model, with some empirical validation, to explain the perception of others in a variety of online scenarios, including short-term and long-term interaction with and without photographs of the participants.

The Role of Photographs

The conventional wisdom in online dating holds that you must attach a photograph of yourself if you want to succeed in meeting someone. In the analysis in Chapter 3, I will explore the effects of photographs on one online dating system. Here, I review prior research on their impact.

Walther et al. (2001) studied online workgroups in a 2x2 design, varying short- or long-term interaction and the presence of photographs. They found that users experienced affection and social attraction:

1. Most of all in *long-term* online groups *without* photographs.

2. Less so in *long-term* online groups *with* photographs and *short-term* online groups *with* photographs.

3. Least of all in *short-term* online groups *without* photographs.

Photographs had opposite effects in short-term and long-term groups. In short-term groups, the presence of photos yielded greater affinity. Walther et al. suggest that photos in this context provide a feeling of social presence that mitigates the asocial tendencies that arise in pseudonymous groups without an expectation of future interaction. The photo's "dose of reality" makes users behave more like they would in a short-term face-to-face interaction (Walther et al. 2001).

In long-term groups, however, the presence of photos reduced affinity. Walther et al. (2001) explain this finding as a result of the hyperpersonal effect — in long-term interaction, users have the opportunity to construct idealized perceptions of others, but photographs thwart the process by injecting a reality opposed to the ideal image.

Similarly, Farnham and Riegelsberger (2004) found that online gaming profiles with photos were preferred less than those without photos; they also found, however, that users formed "more complete impressions" of others when photos were included. Farnham and Riegelsberger (2004) note that we must consider as social engineers whether we want to facilitate idealized perception without photos or promote accurate perception, and perhaps ultimately more successful matching, with photos. Although users might say they are more satisfied with the former, they might be better served by the latter.

TRAJECTORIES OF DISCLOSURE

Progressive self-disclosure constitutes a major part of the development of a romantic relationship. The mystery presented by a new friend or lover enhances their appeal, and gradual disclosure builds intimacy even as the rush of novelty diminishes.

Online personals systems provide a wealth of information at the outset, perhaps eliminating the opportunity for delight at discovering shared musical tastes or favorite books. On the other hand, much of the information, as we have seen, might be exaggerated or falsified, so perhaps users do not have as much information as they think they do.

After users identify someone they like, they can contact the person through a private messaging system. Typically, these systems function like simple Web-based email clients. This is the venue in which much of the initial self-disclosure occurs in online dating. Popular advice columns and books about online dating often suggest

Some newer systems, such as eHarmony, deliberately foster self-disclosure that the site's architects think will facilitate interpersonal affinity. In the initial stages of private messaging, the system does not allow users to send arbitrary text messages, instead asking them to select pre-written questions to that seem intended to encourage

mutual self-revelation, such as:

> If you were taken by your date to a party where you knew no one, how would you respond?
>
>> a) stay close to my date, letting him/her introduce me
>>
>> b) find a spot at the back bar and relax alone, letting him/her work the room
>>
>> c) strike out on my own, introducing myself and making friends
>>
>> d) I would ask my partner if I could skip this particular event
>
> How often do you lose your temper?
>
>> a) practically never
>>
>> b) once in a while
>>
>> c) on occasion during the week
>>
>> d) probably once a day on average
>
> *(eHarmony.com)*

AMBIGUITY AND IMAGINATION

We always contend with incomplete information about others. In many cases, a dearth of information can lead to greater interpersonal affinity than would complete information. Norton et al. (2004) argue that ambiguous characterization allows others to project the attributes they prefer onto a person, leading to a more favorable perception of them. Consider common statements in personals advertisements, such as:

> — I really enjoy good music.
>
> — I like going to movies.
>
> — I enjoy spending time with my friends.

These statements are true about so many people that they provide little or no information, yet they allow the reader to imagine that the movies and music that the writer enjoys match the reader's preferences in those domains. This sort of optimistic interpretation of others relates to the human tendencies to assume that in-group others are like one's self (social projection [Clement & Krueger 2002]) and generally to like people (person-positivity [Sears 1983]).

As evidence for their hypothesis, Norton et al. (2004) conducted a series of studies. First, they found that people *thought* they would like someone better given more information about them. Second, they showed that, contrary to the intuition demonstrated in the first study, having more information was associated with less liking. Third, they found that with repeated trials, the trend reversed, and in the final of four trials, participants liked people better given more information about them.

Intimate Perception

Imagination facilitates the intimate perception of another person, posits Armstrong (2002):

> It's not necessarily the case that the imaginative person can see elements other people are unaware of, it's that they think up less expected — and perhaps more revealing — ways of putting together the elements which anyone can observe. This matters when it comes to love, especially falling in love, in a crucial way. Most people are not startlingly beautiful or magically attractive. But someone who seems just moderately nice — to most people — can flower under the imaginative attention of a lover's eye. Not ... because the lover is somehow gilding the other with fictitious charms; but because the kind of attention the lover brings allows less obvious qualities to be seen and appreciated. Just as a muted work of art ... would quickly be passed over by someone alert to only the most obvious signs of artistic bravura, so a muted person (an ordinary person) has attractive qualities which will probably not be evident to a casual observer. In other words, imagination can be allied to acuteness of perception, rather than to distortion.

> (Armstrong 2002)

What is the role of imagination in online dating? Can people perceive others with sufficient acuity to appreciate nuances of personality? When it comes to the imaginative perception of real qualities, the limitations of multiple choice, short answer, and static photographs would seem to obscure the kind of detail that would "not be evident to a casual observer," which Armstrong suggests that lovers embrace.

Even richer media, like video personals, do not convey the subtleties of personality and mannerism that five minutes of face-to-face conversation would provide. It would seem that most mediated communication provides insufficient detail for the imaginative perception of a romantic prospect to occur in the sense that Armstrong describes.

SELLING YOURSELF (OR BUYING A MATE)

Becker (1992) conceptualizes people's search for mates in terms of a marriage market, in which the participants seek partners to maximize their own happiness or welfare by exchanging what they have for what they want (in Becker's words: to "maximize their expected well-being"). Each person carries a certain value on the market, which we can consider to be a weighted average of characteristics like beauty, intelligence, social status, wealth, and fertility. Logically, people should seek the best partner they can afford; we would all like to have someone who is a "10," but a "7" can realistically expect only another "7." Not all 7's possess the same mix of qualities. One might make up for lower attractiveness with a greater intellect; one "shops" for the preferred combination of qualities. Indeed, single people who use "formal marriage market intermediaries," such as personal ads or human match-makers, sometimes refer to the process of "buy[ing]-a-boy" or selling oneself as "the right product" (Ahuvia & Adelman 1992).

Market models often assume that actors have complete or nearly complete information about the world. Usually, in social environments where one might find a date, this is not true. But online personals systems would seem to provide an ideal example of a marriage market, because they offer so much information about so many choices. (As

we noted above, though, it's unclear whether a computer-based profile can capture enough salient information to gauge attraction or compatibility.) As such, online personals provide a powerful empirical test-bed for market theories, and the theories themselves provide a framework within which we can consider personals systems. (Marriage markets give us little insight into the actions of individuals, who likely are not calculating the marginal benefit of one mate over another, but they are more useful for describing trends across thousands of users of online personals.)

Catalogs of People

In mid-20th century America, the Sears Roebuck catalog provided almost every non-perishable good a person could want, all shipped to your doorstep. Online dating provides such a catalog of potential mates — or dates, or sex partners. From the perspective of the individual, the prospectives seem bountiful; this surplus surpasses the scale of a singles bar by several orders of magnitude.

However, such a bounty demands a different cognitive approach than one would employ at a dinner party, for example. An intimate gathering allows one to discover others gradually, through social and environmental context and through conversation. Of course, we cannot help forming initial impressions based on superficial characteristics, but the organization and scale of such an event allow us to look past them. In online dating, users typically search and sort by relatively superficial characteristics, precluding interaction with anyone who does not meet the criteria the searcher specifies. Browsing a large catalog requires exclusion of entire categories, snap judgments, and quick dismissal of the vast majority of the items. It is unclear to what degree this mentality might affect decisions in online dating.

The Tyranny of Choice

Iyengar and Lepper (2000) studied consumer behavior by varying the number of gourmet jams available at a supermarket tasting table. They found that customers were more likely to buy jam when they were presented with fewer choices at the tasting table; they also liked the jam better. The authors suggest that customers might have enjoyed their choices more when the options seemed more special. Furthermore, they indicate that it might not be dissatisfaction with what they choose but rather the increased uncertainty that comes with choosing among many similar items that causes stress in decision-making, an effect which would be exacerbated if the decision were important and if some of the options were clearly not as good as others.

In a subsequent study, Botti and Iyengar (2003) found that, contrary to popular wisdom about the pleasure of making one's own decisions, those who were permitted to choose were more satisfied than those whose choices were made for them only when selecting from "more preferred alternatives." When the options included "less preferred alternatives," non-choosers were more satisfied (Botti and Iyengar 2003).

It is possible that similar effects color the perceptions of users of online dating systems when they face choices between thousands of potential partners — in the offline world, people never have so many simultaneous dating options. Studies like these also have implications for understanding arranged marriages, in which partners often maintain higher levels of satisfaction than in freely chosen love marriages (Brehm et al. 2002).

Additionally, the perpetual availability of a catalog of others lowers the cost of leaving an existing relationship. Usually, the prospect of being alone for a period of time and the uncertainty of finding another partner serve as disincentives to the termination of a present relationship. With a catalog of single others readily available, people who are advertising their availability to people like you, it should be easier to dismiss the current relationship with confidence that one can find another person at will.

Similarly, online personals make the opportunity cost of entering a relationship — or staying in one — readily apparent. Instead of a vague knowledge that you might be missing out on someone better, you have detailed profiles of all the possibilities.

SOCIAL CONTEXT

In the offline world, we meet friends and lovers in the context of existing social structures. The grocery store, the bar, church, or the neighborhood are common venues for meeting people. More abstractly, one's social network serves as the backdrop for introductions to friends of friends. Interactions which emerge from these contexts remain socially *situated* or *embedded* (Edmonds 1998) within them, such that the involved people are accountable not just to each other but to the wider circle of friends and acquaintances. The sanctions for misbehavior therefore extend beyond the immediate dyad. In terms of dating, this provides incentive for individuals to treat their dating partners well; additionally, it offers some assurance that others are unlikely to behave badly.

Online personals, on the other hand, eliminate most social context. We might consider the online dating site itself a social context, but most such sites have no provision for social interaction outside the goal-directed search and communication activities intended for meeting potential partners. Other kinds of interaction, which might lead to the formation of broader-purpose bonds, are not facilitated. Without a social context in which interactions might be embedded, users can misbehave with fewer consequences than in an offline dating milieu; word of their misbehavior will likely never reach their offline friends, and the unfortunate dating partner can be wiped from the email record and blocked from future communication with a few clicks.

It is also true, however, that online personals cut across offline social contexts. These systems introduce users who would never meet through face-to-face channels because their social contexts do not overlap; indeed, online dating bridges social contexts between which there was no previous connection. In this capability lies the promise of online personals, the real advantage they can offer over face-to-face meetings — a tremendous pool of potential dates, not only large but also distinct from those you would meet any other way.

For certain marginalized populations, the circumnavigation of one's offline social contexts provides the additional benefit of privacy. Gays and lesbians who do not disclose their sexual orientation, for example, might be unable to seek a partner through offline channels without revealing their orientation to friends, family, or co-workers.

Social networking systems like Orkut and Friendster, on the other hand, exploit existing offline social contexts to allow users to meet, e.g., friends of friends. These tools mimic real-world mechanisms for meeting others, with

many of the same advantages and disadvantages. Friendster and Orkut indicate others to whom you are connected by a chain of friends (i.e., "friend of a friend of a friend of a friend"), but distant connections — more than approximately three degrees from you — are effectively strangers.

Narrow-Purpose and Broad-Purpose Communities

Most traditional online dating sites facilitate narrow-purpose community. They offer tools for finding people to date and communicating with them, but they tend not to provide tools for communicating in a broader context or establishing ties outside of a dating context. In addition, their marketing and on-site presentation of communication tools strongly suggest that the purpose of these sites is dating only, even if its tools could be used for other forms of interaction. In particular, the positioning of the private messaging system as a mechanism for contacting potential romantic partners likely discourages same-sex communication on heterosexual systems, even where it might facilitate friendship or prove otherwise beneficial.

By contrast, social network systems like Friendster.com facilitate romantic interaction, but they situate it in a broader context communication — on Friendster, users can message friends of friends for dating purposes, but they can also announce a party to first- and second-degree members of their social network, or locate old friends from high school. Although Friendster has a "suggest a match" function, with which users can introduce two friends who might like each other, it has no dating-specific message features. Because romantic messages travel through the same channels as non-romantic messages, we might consider Friendster a broad-purpose system. Other broad-purpose systems include "portal" sites like Yahoo!.

By providing incentive to interact on the site even when one is not seeking romance, broad-purpose sites may be able to retain their users even while they are off the dating market.

DIFFERENCES IN ONLINE DATING SYSTEMS

Online dating systems take a variety of approaches to the goal of enabling compatible users to meet and date. In this section, I attempt to provide a summary of the different ways these systems handle the key tasks in online dating. These classifications come from several years of observation of dozens of online dating systems, but they are not based on any formal method of taxonomy.

Stages of Online Dating

We can divide the steps that users take as they use online dating systems into three major tasks: describing themselves, discovering other people, and communicating with selected others. Because dating systems differ within these tasks, but almost universally have the three tasks in common, I will describe approaches to the tasks separately rather than trying to categorize sites holistically. I will also give examples of sites that employ each approach; some sites offer more than one approach for the same task. These lists are meant to be representative, using well-known sites, not exhaustive. As with the Web in general, there are too many online dating sites opening and closing every day to craft a comprehensive list.

Self-description

■ *DIRECT*

Most online dating systems construct personal profiles by asking users to describe themselves directly. They request demographic and personal information, including age, sex, sexual orientation, location, marital status, children at present, child-bearing aspirations, race, religion, height, weight, body type, attractiveness, smoking habits, drinking habits, drug habits, and (on at least one system) self-deprecation habits. Even sites like eHarmony that focus on indirect self-description include at least basic demographic and personal questions.

It is in the multiple-choice or short-answer personal and demographic questions that sites for various sub-populations tend to display their specialization. For example, Manhunt.net, a site aimed at gay men, allows users to specify their body type with terms like "bear," which is unique to gay male culture. JDate, a site for Jewish singles, allows them to specify their particular sect of Judaism. Kissykat.com, for pet lovers, gives users the ability to list their pets (and search for others based on the pets they own).

Sites that employ primarily direct description also prompt users in direct language to describe themselves and what kind of person they want to meet (e.g., "The most important thing in my life is…").

Examples: Match.com, Yahoo! Personals, Spring Street Networks (provider for Nerve.com, Salon.com, etc.), many others

■ *INDIRECT*

Social psychologists routinely find that people are unable to describe themselves accurately on a variety of dimensions, even when they believe they can. Perhaps because of this, some sites have begun to provide indirect ways to prompt users to describe themselves. Nerve.com includes profile questions that allow users to self-disclose

Figure 1.2. Physical attractiveness test from Match.com. The screen on the left asks the user, "Pick your favorite or the one **you could tolerate best.**" At right, the screen instructs, "Click on photos you find personally **attractive, cute, or sexy.**" (Highlights theirs.)

without having to make explicit statements about themselves. Questions like this on Nerve include "Song or album that gets you in the mood," which brings out musical preferences as well as providing an opportunity for innuendo; "Five things you'll find in your bedroom," which gives users a chance to mention personal details and, again, allude to sex; and "Celebrity you resemble most," which allows users to provide a reference point for appearance, sometimes ironic, and to demonstrate affinity (if any) for pop culture.

More recently, sites like eHarmony.com and Tickle.com administer personality tests as a major part of the self-description process. These tests purport to gauge how users behave in social situations, how they handle conflict, and what is important to them in romance — but the sites provide no particular evidence of what their tests measure, so it is unclear how well they correspond with the personality psychology literature. However, research has shown that personality factors like communicative style and conflict management as well as emotional disposition are important to successful relationships (Brehm et al. 2002).

Match.com developed a physical attraction test (Figure 1.2) in conjunction with weAttract.com. As the eHarmony and Tickle tests assess personality without asking the user to report it explicitly, so the Match.com test allows users to express their preferences in physical characteristics by demonstrating them in a series of simple choices. The test presents a series of rating and comparison tasks. In the rating tasks, users indicates on a Likert-type scale how much he or she likes an image of a person's face or body. (The images show models, but they are not all model-beautiful.) In the comparison tasks, users are presented with two or more images of faces or bodies and must choose the one they like best (or dislike least). Using these responses, the system constructs a model of the user's preferences for various physical attributes.

Examples: eHarmony.com, Tickle.com, Match.com, Spring Street Networks

■ *FREEFORM*

A few online dating systems, rather than creating elaborate profile systems, simply extend the concept behind newspaper personal ads, which are usually a few short lines of text printed as classified ads. These systems give users a name or headline and a space for whatever text the user wishes to include. Of course, these ads can be much longer than newspaper personals, for which people typically pay by the word or line, but in the same spirit, they permit unstructured text.

Examples: MakeOutClub.com, CraigsList.org

Discovery of Others

■ *SEARCHING*

The most basic way to discover other people in an online dating system is searching. At the basic level, this allows users to specify the sex, age, and location they seek; most sites provide more powerful capabilities as well, so users can search for, e.g., all the brown-haired, blue-eyed men between 5'11" and 6'3" who don't drink within a five-mile radius of zip code 90210.

Searching gives users a great deal of control, but sifting through so many options can be overwhelming. Advanced searches also give users the ability to overspecify, so that searches return only a few people who meet very specific criteria. This is problematic if people poorly understand who might be compatible with them; in offline interactions, one is not able to filter the people one encounters so precisely.

Notably, eHarmony.com, which relies heavily on personality-based matching, described below, does *not* include search functionality. That is, users cannot conduct searches for people who match arbitrary sets of characteristics; they can see only the users with whom the system deems them compatible.

Examples: Match.com, Yahoo! Personals, Spring Street Networks, many others

■ *MATCHING*

Many online dating systems will automatically match users with others whom the system deems compatible by comparing their profiles. Companies keep their matching algorithms private, but some provide limited information about them to users.

Some matching algorithms simply look for similarity. One common method for computing a similarity score involves representing a profile with n features as a vector in n-space whose similarity to another profile can be conceptualized as the degree to which the vectors point in the same direction. Others also check for mutual satisfaction of requirements — is person A's age in the range person B seeks, and vice versa?

Sites that employ personality tests, like eHarmony and Tickle, focus on complementarity more than similarity. Although their algorithms are carefully kept secret, they claim to pair people with personality traits known to complement each other in successful relationships. Without knowing the details of their matching processes, it is difficult to assess how accurately they do this.

Typically, sites present users with a list of their matches along with, in some cases, a "compatibility score" that indicates the relative strength of each match. As mentioned above, eHarmony.com is apparently unique in revealing to users only those people with whom the system considers them compatible — users cannot view any other profiles.

■ *SERENDIPITY*

Finally, some systems include mechanisms for exposing users to randomly selected others, or allowing them to stumble across profiles without having to search for them or be matched with them. Usually, this takes the form of a "user of the day" profile, featured on the home page, or a different randomly selected profile every time the home page is loaded.

Examples: Spring Street Networks, MakeOutClub.com

Communication

■ *PRIVATE MESSAGING*

Most online dating systems provide private messaging systems, essentially Web-based email that uses pseudonymous handles instead of real names and email addresses in order to protect the privacy of users until they choose to reveal it to their correspondents on the site. Commercial systems often focus their business models around the private messaging system, charging either a monthly fee or a per-message fee to send mail. Many allow users to receive messages for free.

Match.com also lets users send private voice messages.

Examples: Match.com, Yahoo! Personals, Spring Street Networks, many others

■ *ATTENTIONAL TOKENS*

Some sites offer a way to express interest in someone without sending a full message. The names of these tools sometimes evoke physical metaphors — "eye contact" or "wink." One site refers to them as "collect calls." They have in common that the action is free and notifies the recipient that the sender has expressed interest in him or her, but they typically include either no textual content or one of a few prewritten statements from which the user can choose.

Examples: Spring Street Networks, Match.com

ANALYSIS OF BEHAVIOR ON ONE DATING SYSTEM

This introduction has provided a survey of diverse perspectives on attraction and online dating. In the second chapter, I present a quantitative analysis of the behavior of users on one online personals Web site. The analysis seeks to reveal the preferences of the users with respect to potential dating partners by examining whom they contact and with whom they converse from a set of 57,000 active users.

The findings highlight the importance of a variety of characteristics to users of the site under study, as revealed by the communicative patterns of the users. In particular, the examination of "bounding" characteristics shows that life course attributes, including marital status, whether one wants children, and how many children one has already, are much more likely than chance to be the same across the two users in a dyadic interaction. Regression models built to predict how many messages a user with specific attributes will receive reveal somewhat different preferences for men and women. For both sexes, the strongest predictor by far of messages received is the number of messages sent. For men, age, educational level, and self-rated physical attractiveness are the next most important qualities. For women, however, they are not being overweight, self-rated physical attractiveness, and having a photo.

In the third chapter, I consider the implications of these findings for the design of online dating systems. Specifically, I suggest elevating the importance of life course attributes in both search interfaces and matching algorithms to better reflect their great importance to users. Additionally, I note that many subsets of the population

— by sex, ethnicity, religion — might benefit from interfaces tailored to their unique preferences; a generalized tool that serves all subpopulations likely serves all of them suboptimally.

Finally, I propose new directions both for the design of online dating systems and for their further study. New designs might include mobile, in-context systems; experiential tools for learning more about others; and the incorporation of physiological signals of attraction. My research agenda suggests further analytic work, but also experimental, ethnographic, and comparative approaches.

Chapter 2

Analysis of Users and Interactions

GOALS OF THIS WORK

In taking an analytic approach to understanding how people use online dating systems, I have two primary goals:

1. Gain a better understanding of which factors in human attraction are salient in an online personals context

2. Inform the design of future online dating systems by providing them with information about what factors people actually consider when they make decisions about whom to contact and answer

To achieve these goals, I analyzed data from one online dating system in particular. Through an agreement brokered by the Media Laboratory with an online dating Web site (the "Site"), I obtained access to a snapshot of activity on the Site over an eight-month period, from June 2002 through February 2003. The data included users' personal profile information, their self-reported preferences for a mate, and their communications via the site's private message system with other users.

ANALYTIC APPROACH

We received the database tables as comma-delimited text files. Through a series of Perl and SQL scripts, these files were transformed into PostgreSQL relational tables on our database server. This server provided data for statistical analysis and interactive visualization. Over the course of this project, I added tables to the database that contained summary statistics and aggregations of user behavior into dyadic conversations instead of individual messages. Most summary statistics were normalized by the number of days each user had been a member of the site; the metric *number of messages per day* is not confounded with membership duration, as the raw metric *number of messages* would be.

Statistical Packages

To understand the patterns of interaction in the data, I used two statistical packages, SPSS 11.0 and the open-source R 1.7.1 (a clone of the commercial statistical environment S). With these tools, I generated descriptive, comparative, and predictive statistics about the attributes and behavior of individuals and dyads.

Visualization Tool

Additionally, I built a flexible interactive tool to generate visual representations of the users in the data set and the connections between them. I built this tool using the Cocoa libraries on Mac OS X to download data directly from the PostgreSQL database and display it using OpenGL. It supports plotting and coloring users according to a variety of numerical and categorical characteristics as well as overlaying the network dyadic ties among the users. The visualization can be reconfigured on the fly; it is essentially a researcher's interface for exploring this data set.

Rajiv Eranki, an MIT undergraduate who undertook a semester of research with the Sociable Media Group, also made substantial contributions to this tool.

NATURE OF ONLINE PERSONALS DATA

Table 2.1 indicates which profile characteristics users could specify about themselves and about the partners they would like to meet. Appendix A lists the possible values for the categorical attributes.

Data about private messages exchanged by the users included the sender, recipient, subject, text, date and time of delivery, and whether the recipient had read the message.

The Site's data also included information about the most recent search for profiles that each user had conducted, if any.

Definitions

A few definitions will allow for clearer, more concise expression of the findings in this section.

Contact or *conversation* — A series of one or more private messages exchanged between two users during the studied period.

Initiator — The user who begins a conversation by contacting another user with an initial private message. Sometimes called "Person A."

Recipient — The user who receives an initial private message from another user and sometimes responds. Sometimes called "Person B."

Attribute	Type	Self	Partner
Username (online handle)	Free	■	
Gender	Cat	■	■
Birth date or age	Con	■	■
Height	Con	■	■
City, state, and postal code	Cat/Free	■	■
Build	Cat	■	■
Drinking habits	Cat	■	■
Smoking habits	Cat	■	■
Education level	Cat	■	■
How many children user has	Buck	■	■
How many children user wants	Buck	■	■
Martial status	Cat	■	■
Pets owned	Cat	■	■
Pets preferred	Cat	■	■
Attractiveness	Buck	■	■
Race	Cat	■	■
Relationship type sought	Cat	■	■
Religion	Cat	■	■
Importance of age in a partner	Buck	■	
Importance of height in a partner	Buck	■	
Text self-description	Free	■	
Text preferences for partner	Free	■	

Con = continuous   **Cat = categorical**
Buck = bucketed scale   **Free = free response**

Table 2.1. Profile attributes specified by users about themselves and about potential partners

Figure 2.1. At top, geographic distribution of users of the Site under study. Users, each represented by a point, are plotted with the visualization tool developed for this project by latitude and longitude, resulting in a projection of the geography of the United States. Intensity of color indicates population density. At bottom, an overlay is added showing the communications between users. Brighter paths indicate communication between those locations by multiple pairs of users. Broader paths indicate greater numbers of messages exchanged by a dyad. This view shows regional and trans-national ties, but it is important to note that most communications take place across a distance of less than 20 miles. These visualizations were created using the visualization tool built for this project.

GEOGRAPHIC DISTRIBUTION

Although the Site has a national base of users, they are distributed differently from the U.S. population on a state-by-state basis (Table A.1 in Appendix A). Heavy use appears in upstate New York, the Southeast, the Midwest and Great Lakes regions, and certain secondary urban areas in the West, such as Sacramento, Calif. Figure 2.1 shows a plot of users by geography, with communication between them overlaid. Although this display highlights long-distance communication, most contact occurred over short distances (Figure B.7 in Appendix B).

Two factors may account for this irregular distribution, and perpetuate it: first, advertising by the company seems to target secondary urban and suburban markets rather than major coastal markets like New York and Los Angeles. Second, because online dating systems require a critical mass of users — i.e., an adequate catalog of people — to be useful, areas that reach critical mass will grow even faster, because their selection of potential mates continues to become more appealing, while areas without enough users will have comparatively more trouble recruiting new users. (Grounded in geography by the desire of users to meet others in person, online dating systems require critical mass in each area separately.)

USER DEMOGRAPHICS

The Site had 221,800 members as of February 2003, the end of the eight-month period covered by our snapshot of the Site's data. Of these, 69.4 percent (153,942 users) had fully completed their profiles. A slightly different subset, 25.9 percent of the total (57,362 users), was active during the eight-month study period — that is, they sent or received at least one message. This active subset was used for analyses of messaging behavior, but for analyses involving profile characteristics, it was trimmed to the 23.8 percent of users (52,857) who were active and had complete profiles.

Although the overall user population on the Site included more men (62.8 percent) than women (37.2 percent), the active subset analyzed in this work was 55 percent female. The Site targets heterosexual users; although it allows users to specify same-sex preferences (e.g., "male seeking male"), less than one percent of users did so, and some of these appeared to be data entry mistakes or confusion in the interface about whether, for example, one *is* male or *is seeking* a male. Because homosexual users were so few, their behavior would be inadequate to draw conclusions about gay users' behavior in online dating environments; thus, these users were excluded from the analysis.

Within the active subset of users, the median age was 34, but the male population was slightly older (median 36 years, compared to 33 years for women; see Figure A.2). Most users were Caucasian (83.7 percent); African-Americans and Hispanics each composed approximately two percent of the user population (Table A.4). Nearly 10 percent chose not to give their race.

The majority of active users were Christian (54.9 percent). Agnostics were the next largest group, with 1.8 percent of users; 0.9 percent were Jewish. Most of the remaining users chose the "Other" option (7.2 percent) or preferred not to answer (34.2 percent); small numbers were Muslim, Hindu, or Atheist (Table A.5).

Users were almost evenly split between never having been married (37.7 percent) and having been divorced (39.4

percent). Some were separated (5.6 percent), and small numbers were in a relationship, married, or widowed (Table A.6).

Many users reported having at least one child (42.0 percent; Table A.10). A question about whether the user wants children was ambiguous about whether the children would include existing children or be in addition to them, and 30 percent of respondents said they would prefer not to answer (Table A.11). Of these, about a third wanted children, a third did not want children, and a third were undecided.

High school was the highest educational level attained by 15.3 percent of users (Table A.7). A plurality, 36.8 percent, attended some college, and 23.4 percent held a college degree. Approximately 11 percent reported at least some post-college education.

Most users were non-smokers (60.4 percent); 21.6 percent reported social or regular smoking, and 7.3 were trying to quit (Table A.9). Social drinkers constituted 76.8 percent of the population, compared to 11.5 percent non-drinkers and only 1.8 percent regular drinkers (Table A.8).

A plurality of both men and women described their build as average (36.7 percent). Among women, 23.9 percent said they were petite or slender, compared to only 6.8 percent of men (t = -52.468; d.f. = 57,343; p < 0.001). Conversely, 33.6 percent of men reported that they were athletic or "body-builder," versus 9.3 percent of women (t = 73.501; d.f. = 57,343; p < 0.001). Approximately 10 percent of both sexes said they carried a few extra pounds (Table A.12).

On a three-point scale of Average, Attractive, and Very Attractive, most users (54.2 percent) said they were attractive (Table A.13). Women were nearly twice as likely as men to say they were very attractive (t = -19.274; d.f. = 57,343; p < 0.001), but men were about 1.5 times as likely as women to describe themselves as average (t = 26.529; d.f. = 57,343; p < 0.001). For the middle rating, Attractive, the sexes chose more similarly. Women were only 1.08 times as likely as men to choose this answer — less than 10 percent more — though the difference was also significant (t = -4.813; d.f. = 57,343; p < 0.001).

MESSAGES AND CONVERSATIONS

During the eight-month period from June 2002 to February 2003, 29,687 users sent 236,930 messages to 51,348 users.

In total, these messages constituted 110,722 exchanges of one or more messages between unique pairs of users, which I term a *conversation*. However, most of these dyadic exchanges were something less than dyadic: 78.2 percent (86,597) of conversations consisted of unreciprocated single messages.

Messages were received in a more even distribution than they were sent; that is, fewer members sent messages than received them. Users sent and received a mean of 1.50 messages (median = 0.0) in the eight-month study period. (The means are the same because the same bounded population sent and received the messages.) However, the standard deviation for number of messages sent was 7.45, as compared to 4.90 for number of messages received,

indicating that messages were distributed more evenly across the set of recipients than they were across the set of senders. In total, 29,687 users sent 236,930 messages to 51,348 users.

Same-sex messaging

Nearly all conversations (97.0 percent) were between a man and a woman, but there were 3,270 exchanges between two men or two women. In these conversations, a few users expressed same-sex romantic or sexual intent, or they were looking for same-sex friends. However, most such messages were due to one of the following:

> **User error in entering gender.** Some users mistakenly appear to have selected the wrong sex when they filled out their profiles, either by mechanical error or by confusion over which sex they *were* versus which they were *seeking*. These users were identified anecdotally by messages they sent that indicated a sex other than the one they specified; the prevalence of this error is unknown.

> **Test messages sent by staff of the Site.** In testing the functionality of the Site, members of its staff sent a number of test messages to each other. Some of these exchanges were same-sex.

> **Announcements sent by staff of the Site.** A male employee of the Site periodically sent service announcements through the private messaging system to some or all users, male and female.

> **Spam.** A few users sent a large number of junk messages through the private messaging system. The "Nigerian" scam was most common.

> **Account sharing.** Some users specifically mentioned in their messages that they were writing from the Site account of an opposite-sex other (e.g., a woman writing from her brother's account) so that they would not have to pay for a membership. This behavior is contingent on the pricing model used by the Site and many others, in which users receive unlimited message-sending capability for a flat monthly fee. It would likely occur less on a site with a per-message pricing model.

> **Writing on behalf of another.** Some women wrote from their own accounts to other women on behalf of another, that is, providing a third-party introduction (e.g., "My brother is a great guy...").

> **Backchannel communication.** A very small number of women communicated with other women about men on the Site. In the eight-month sample studied here, no men communicated with other men in this manner. One example from a woman to a woman:

>> Hi! I was checking what my horoscope said on my profile, and our pictures where [sic] next to each other. I thought, "I wonder if any other women are having any luck with this site?" I guess, technically, we are supposed to be in competition with each other, but why not compare notes? :)

Because the sample of same-sex contacts was small and furthermore included few typical personals interactions —

(a) Men **(b) Women**

Age (years)   Age (years)

Mean ties initiated
Mean ties received
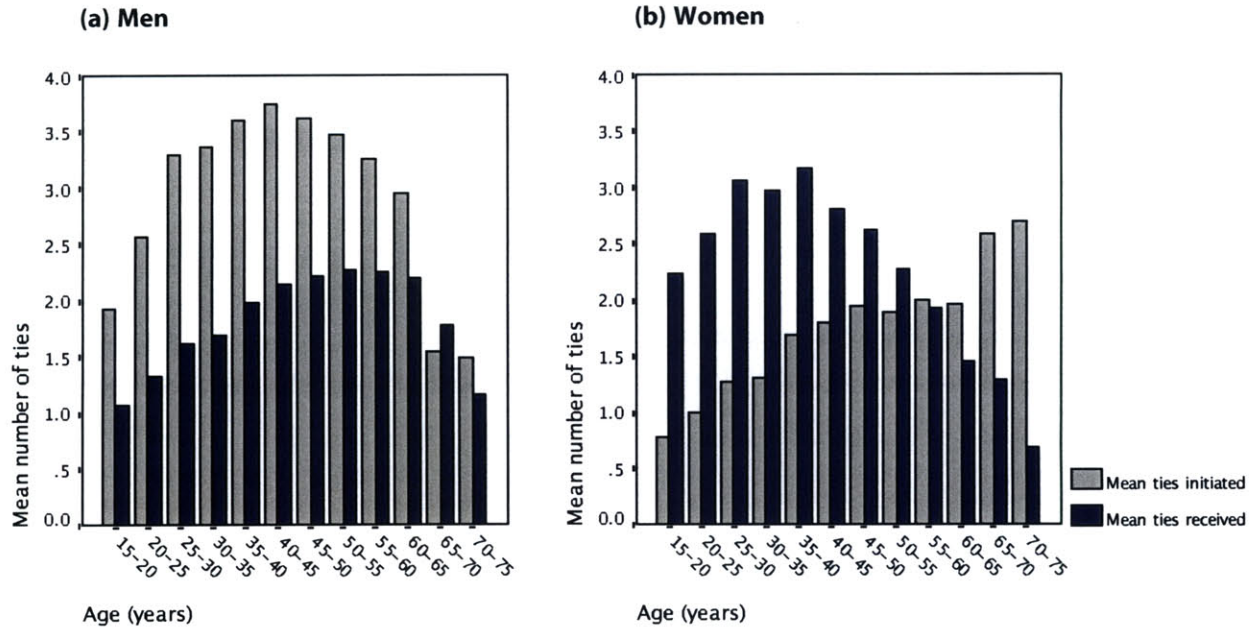
Figure 2.2. Number of dyadic ties initiated and received (based on first message), by age and gender, in different-sex interactions.

that is, most interactions had purposes other than meeting people to date — I excluded same-sex interactions from these analyses. A site directed at a same-sex population, or at least including a significant gay or lesbian population, would be more useful than the present Site for understanding same-sex preferences and behaviors in online dating.

Different-sex messaging

Of the different-sex exchanges, men initiated the majority of conversations (73.3 percent vs. 26.7 percent); however, their initiations were 17.9 percent less likely to be reciprocated than those begun by women (20.6 percent reciprocated vs. 25.1 percent for female-initiated; $t = -15.465$; d.f. = 50,150; $p < 0.001$). Tables B.8 and B.9 in Appendix B provides more detailed statistics about reciprocation, categorizing the exchanges further by whether the participants satisfy each others' age requirements.

Users of both sexes had contact with a median of 2.0 distinct others. The distribution was wider, though, for men than for women (mean = 5.3, s.d. = 11.8 for men; mean = 4.2, s.d. = 5.9 for women). Men participated in more communications on average than women, but we would expect this because the active subset of users

Figure 2.3. Number of dyadic ties per person.

Number of dyadic ties
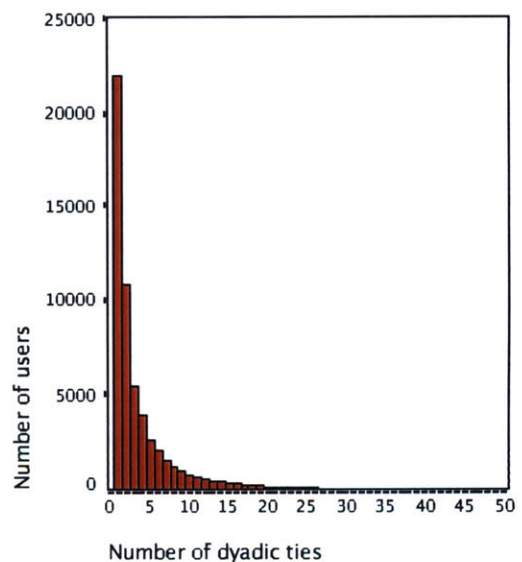
contains more women than men, so the contacts are spread across a larger number of women.

Overall, the number of dyadic ties per person (Figure 2.3) followed the familiar "power law" distribution, with many users with few ties and exponentially fewer with many ties. The mean number of ties per person was 5.0 (s.d. = 9.04).

As expected from the above finding that men begin most conversations, men on average initiated more contacts than they received (mean = 3.3, median = 1.0, s.d. = 7.1 initiated vs. mean = 1.9, median = 1.0, s.d. = 2.8 received). Women, on the other hand, initiated fewer contacts than they received (mean = 1.5, median = 0.0, s.d. = 3.4 initiated vs. mean = 2.7, median = 1.0, s.d. = 3.5 received). Figure 2.2 shows contacts initiated and received by age as well as gender.

From this point forward, the discussion will address different-sex messaging only.

CHARACTERISTICS OF DYADS

This section describes the characteristics of the different-sex dyads formed through private-messaging contact between pairs of users.

Messages Exchanged

In male-initiated exchanges, the man sent an average of 1.65 messages per exchange and the woman sent 0.44. The averages rise to 2.86 for the man and 2.06 for the woman among the exchanges in which she reciprocates at least once.

In female-initiated exchanges — rarer, but more often reciprocated — the woman sent an average of 1.81 messages per exchange, and the man sent 0.56, but if we examine only those exchanges in which he reciprocates at least once, the woman averages 2.76 and the man, 2.20.

For both male- and female-initiated exchanges, the median number of messages sent by the initiator was 1.0, and the median number sent by the recipient was 0.0. In reciprocated exchanges, the medians rose to 2.0 and 1.0, respectively.

Of the reciprocated exchanges, those initiated by women lasted a median of 4.0 days; the male-initiated exchanges lasted a median of 3.6 days.

Age

Users specify their preference for the age of a potential partner as a range. To facilitate comparisons with the user's own age, however, I took the midpoint of that range as an approximation of the age they seek. Figure 2.4 shows the distributions for men and women of the difference between the user's age and the age they sought. Women indicated they were seeking men older by a median 2.5 years than themselves (mean = 3.0). Similarly, men said they were looking for women a median 2.5 years younger (mean = 2.3).

In fact,men actually contacted women who were a median 4.0 years younger (mean = 4.2, s.d. = 6.3)., while women initiated contact with men who were a median 2.0 years older (mean = 1.8, s.d. = 5.3). Across all exchanges, the man was a median 3.0 years older than the woman.

So, women overestimated how much older they wanted men to be, contacting younger men than they initially claimed they were seeking. Men, on the other hand, underestimated how much younger they wanted women to be; they contacted younger women than they initially said they were seeking.

Age Fitness Metric

To better assess the impact of age preferences on communication between users, I synthesized a metric that integrates whether a user's age falls into a conversational partner's preferred range with the self-reported importance of age to the partner, which ranges from 1 (unimportant) through 5 (very important). Specifically, if person A's age falls within person B's preferred range, I define A's age fitness with respect to B as 1.0 * B_age_importance, yielding a value between 1 and 5. If person A's age falls outside person B's preferred range, A's age fitness is defined as −1.0 * B_age_importance, yielding a value between −5 and −1.
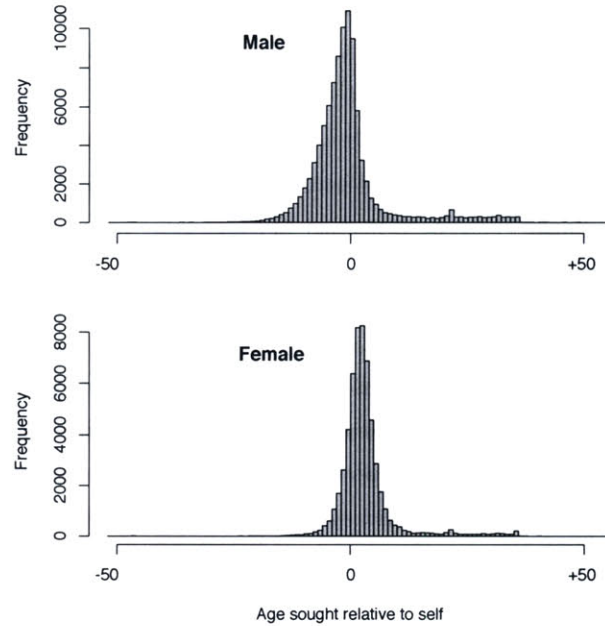
Figure 2.4. Self-reported age sought relative to user's own age, in years.
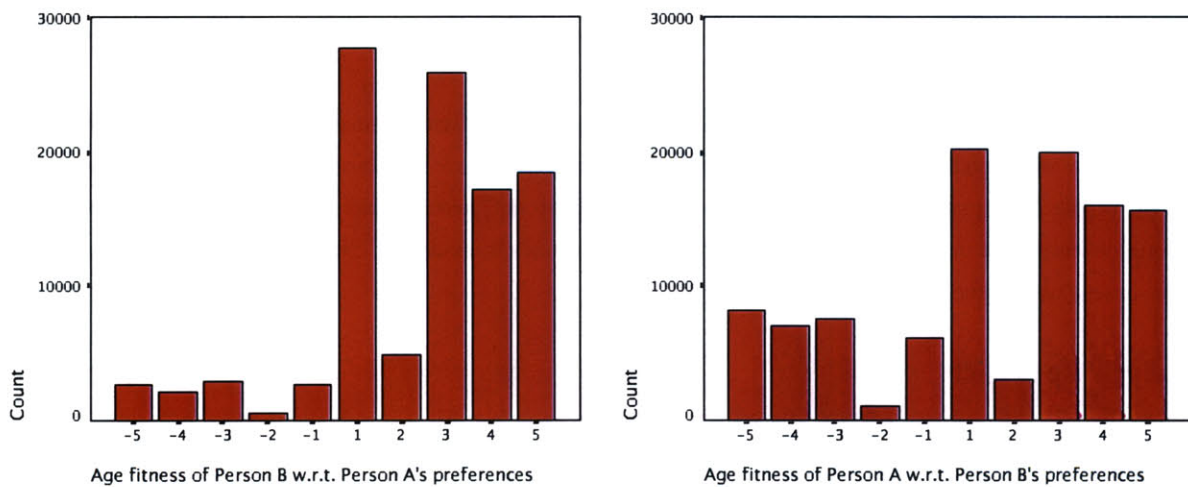
Figure 2.5. At left, a graph of how well the Recipient's age fits the Initiator's preferences. At right, how well the Initiator's age fits the Recipient's preferences. Initiators pay more attention to how well the Recipient's age fits their preferences than to how well they fit the Recipient's preferences. These figures depend entirely on the actions of Initiators, because Recipients cannot control who contacts them.

Users who initiated conversation were naturally more attentive to how well their recipient's age fit their preferences than to how well their age fit their recipient's preferences. Thus, the age fitness of the receiving person (Person B) with respect to the preferences of the initiator (Person A) had a median value of 3 (mean 2.30, s.d. = 2.37), whereas Person A's fitness with respect to Person B's preferences was lower, with a median of 2 (mean 1.23, s.d. = 3.23). This makes sense, because Person B did not choose to be contacted by Person A. Figure 2.4 shows the distributions of fitness scores for initiators and recipients.

BOUNDING CHARACTERISTICS

Some characteristics are more bounding than others — that is, users are more likely to seek someone like themselves on that dimension. For example, smokers might want to find other smokers more so than people with blue eyes want to find other people with blue eyes. Figure 2.5 shows strong bounding on the characteristic *marital status*. The two primary groups are people who have never been married, and people who have gotten divorced or separated. As I will show in the following analysis, people in these groups tend to contact others in their own group.

To determine the bounding strength of the categorical descriptors in the data set, I compared the percentage of contacts between two users who shared the same value for a characteristic (e.g., the value "athletic" for the characteristic "physical build") with the percentage of contacts we would expect to share the value if one male user and one female user from the active user population were paired randomly.

Relationship to Social Network Analysis

This analysis is similar in spirit to studies of segmentation in social networks. However, I chose not to apply traditional social network analysis to this network of users because the nature of a tie in an online personals system differs from that supposed by social network literature. In this environment, ties are dyadic and generally private — users usually do not know anything about the network beyond those to whom they are tied personally. Although the structure that results from messaging activity in online dating systems resembles a social network, its practical significance is somewhat different.

In particular, ties are not transitive: even if A talks with B and B talks with C, it is not likely that A and C know each other or are interested in knowing each other. Assume A and C are male and B is female — A and C did not join the online dating site to communicate with their competitors. These interactions are fundamentally dyadic; although users might maintain several dyadic interactions at once, they do not constitute a typical social network. Thus, in this work I focus on pairwise interactions.

Analytic Approach

If 32.6 percent of male users and 9.2 percent of female users report their build as "athletic," and users were contacting each other randomly but in heterosexual pairs, we would expect 0.326 * .092, or 3.0 percent, of contacts to involve two users of athletic build. However, if users of athletic build sought other such users more often, the percentage of contacts involving two of these users would exceed 3.0 percent; if these users avoided each other, the percentage would be lower.

Figure 2.6. At top, distribution of men and women by marital status. At bottom, communication via private messaging is overlaid. Marital status was the most strongly bounding characteristic, with most communication occurring between two users who share the same status. These images were created with the custom visualization tool. See Appendix B, Figure B.13 for another example.
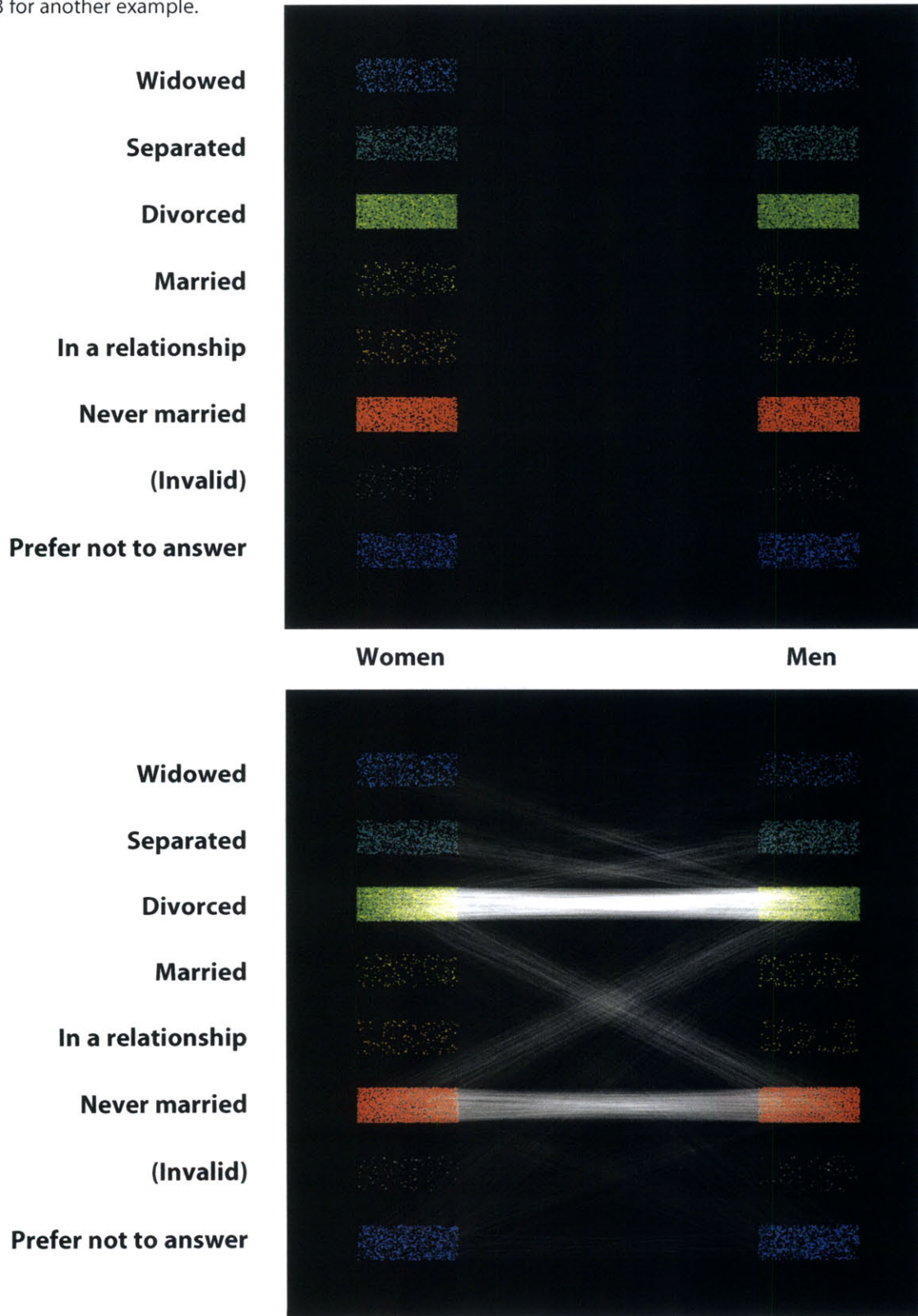
Characteristic	Expected percent same (x)	Actual percent same (all contacts, a_1)	Actual percent same (recip. con. only, a_2)	t ($a_2 \neq x$)
Marital status	31.6	51.7 (1.64x)	56.0 (1.77x)	76.001[†]
Wants children	25.1	38.7 (1.54x)	40.5 (1.61x)	48.553[†]
Num. of children	27.8	38.7 (1.39x)	38.6 (1.39x)	34.352[†]
Physical build	19.2	24.5 (1.28x)	25.6 (1.33x)	22.435[†]
Smoking	40.5	50.6 (1.25x)	54.0 (1.33x)	41.979[†]
Phys. appearance	37.6	46.1 (1.23x)	49.2 (1.31x)	35.886[†]
Educational level	23.6	28.0 (1.19x)	29.3 (1.24x)	19.360[†]
Religion	42.4	49.7 (1.17x)	52.6 (1.24x)	31.589[†]
Race	71.1	81.2 (1.14x)	85.9 (1.21x)	65.808[†]
Drinking habits	61.2	68.7 (1.12x)	73.4 (1.20x)	42.692[†]
Pet preferences	34.7	38.5 (1.11x)	39.9 (1.15x)	16.425[‡]
Pets owned	21.8	23.6 (1.08x)	24.0 (1.10x)	8.038[‡]

[†] d.f. = 23,940; $p < 0.001$ [‡] d.f. = 23,855; $p < 0.001$

Table 2.2. Bounding strength of categorical characteristics. *Expected percent same* indicates the statistically expected percentage of dyadic pairs who share the same value for the listed characteristic. The expected probability is based on random selection from the male and female population distributions for the characteristic. *Actual percent same* indicates the empirical percentage of dyadic pairs who shared the same value for the listed characteristic, across all contacts and just the reciprocated subset, in which the initial recipient replied.

By summing the probability of sameness across all possible values of a characteristic, we find an overall probability that a random pair of one male and one female user will share the same value for that characteristic. These overall probabilities are listed in Table 3.2 as *Expected percent same*. The expected sameness for a characteristic varies with the number of values possible for that characteristic and how evenly users are distributed among the values. Expected sameness is higher when the number of values is low, as with *Physical appearance* ("Very attractive," "Attractive," "Average," "Prefer not to answer"), and when many users have picked the same value for a characteristic, as with *Race* (83.7 percent reported "Caucasian").

Having calculated the expected sameness, I computed the actual percentage of dyads with the same value for each categorical characteristic both for all pairwise exchanges and separately for the subset that were reciprocated. The absolute value of the difference between the actual percentage of sameness and the expected percentage of sameness indicates how much users were deliberately seeking someone with the same value as themselves.

An actual sameness percentage close to its expected sameness percentage indicates that users who share a value for that characteristic did not communicate more often than we would expect by chance if users were contacting each

other randomly. On the other hand, a large difference between actual and expected sameness percentages would indicate that users who share a value for a characteristic communicated more often than we would expect by chance.

Because we expect statistically a varying likelihood of sameness for various characteristics, the absolute difference in expected and actual percentages does not facilitate comparisons between different characteristics, which have different expected percentages. Instead, I calculate the proportion of the actual to the expected percentage sameness for each characteristic. Table 3.2 shows these values in parentheses following the actual percentages for all contacts and for reciprocated contacts. The characteristics are listed in descending order of this proportion, which shows the relative bounding strength of each.

Findings

Users opted for sameness more often than chance would predict in all the characteristics examined in this section. This concurs with the overwhelming evidence gathered by relationship researchers (see surveys in Brehm et al. 2002, Fisher 1992) that actual similarity and perceived similarity in demographics, attitudes, values, and attractiveness correlate with attraction (and, later, relationship satisfaction). However, users demonstrate this homophily to differing degrees for different characteristics.

Dyads were much more likely than chance to choose the same value for characteristics relating to the life course. Values for marital status and wanting children were the same in dyads 64 percent and 54 percent more often, respectively, than would occur with random pairings. The number of children users already have was the same in dyads 39 percent more often than chance. These were the three most strongly bounding characteristics.

Physical build was the same among dyads 28 percent more often than chance would predict. This finding rests on similarity-seeking among a few possible values for build, such as "average" and "athletic," that encompass both genders; many of the other possibilities, like "petite" and "body-builder," are strongly gendered and thus very unlikely to be the same in a heterosexual dyad. Physical appearance — a self-reported rating of attractiveness — was the same among dyads 23 percent more often than chance.

Among lifestyle choices, including smoking habits, drinking habits, and pet preferences, only smoking was the same in dyads more than 20 percent more often than chance would predict. Most dyads (68.7 percent) were the same in drinking habits, but this is because 75.6 percent of male users and 77.9 percent of female users identified themselves as "Social/occasional" drinkers. Thus, the expected probability of sameness was also high for this characteristic, rendering the high actual similarity unremarkable.

Pets — both general preferences regarding them and specific pets already owned — proved the least bounding of any characteristics. Users picked dyadic partners who shared their preferences only about 10 percent more often than chance would predict. Homogeneity on these characteristics did not matter to users nearly as much as other characteristics.

Religion was the same in dyads 17 percent more often than chance. More than half of the active users of the Site

47

identified themselves as Christian, and about a third chose "Prefer not to answer," a very high percentage compared to other characteristics. Given the distribution of religions (Table A.X) among users who did answer, we might reasonably presume that a large number of "Prefer not to answer" respondents are in fact Christians, even if we allow that non-Christians might be more likely to choose "Prefer not to answer."

If this is the case, the bounding strength of religion might appear lower than it is because of users' reluctance to specify their religion. It might also be true that, as some analyses suggest (e.g., Williams & Lawler 2003), having similar religiosity is more important than sharing a specific religion.

The overwhelming majority of dyads (81.2 percent) shared the same race, but, as with drinking habits, this high rate of similarity is only moderately better than chance (14 percent). Because 83.7 percent of users were Caucasian, the rate of similarity expected by chance was also high, 71.1 percent.

In general, characteristics were slightly more bounding among the subset of reciprocated contacts, but the difference was small and roughly equal across characteristics. Although the difference is small, it suggests that users were slightly more likely to respond to an initiation from a more similar other.

MESSAGES EXCHANGED: A PROXY FOR SOCIAL AFFINITY

In this attempt to understand what users consider as they evaluate potential partners, it was necessary to choose a quantitative descriptor from our dataset as a proxy for social affinity. Any descriptor would be less accurate than an explicit rating of a user by a conversational partner, but for this post hoc analysis we were unable to ask users to rate each other. Thus, we chose the number of private messages users exchanged through the Site during our 8-month study period as the dependent variable by which to assess the effect of profile characteristics and user behaviors on the affinity that others had for him or her.

Using the number of private messages exchanged by a dyad as a proxy for social affinity seems reasonable, but it is complicated by the common practice among pairs of interacting users of transitioning to another medium, such as e-mail, IM, telephone, or face-to-face, once the users are comfortable with each other. The sooner the transition to another communicative channel, the fewer messages the pair exchanges through the dating system, which means that the most successful pairings might appear unsuccessful if the number of on-site private messages is the only metric.

In the following section, I analyze the rate of transition to other channels for two reasons: first, to understand its impact on the validity of private message volume as a proxy for social attraction, and second, because the frequency of transition to other channels is important to understanding how users handle the process of getting to know someone through an online dating system.

Transitioning to other channels

The limited scope of the data in this research and the fact that the analysis is post-hoc make it difficult to determine precisely how often users move their conversations from the Site's private messaging system to another

communication channel, such as email or telephone. Without this information, we cannot tell whether a dyadic exchange on the Site has ended because the users have stopped communicating or merely because they are communicating elsewhere. The former indicates an unsuccessful coupling, but the latter might indicate a successful one.

The duration of communication in the private messaging system is the best approximation in this data set of the success of a conversation between two users — so ambiguity about the meaning of an end to messaging between two users clearly weakens findings that depend on the duration statistic, at least until we know the degree to which users transition their conversations to other media.

It is not important that provision of an email address be reciprocated; once one party has contacted the other via email, reciprocal communication is possible even if the latter party did not provide an email address via the Site's messaging system. Thus, a dyad may transition to email as soon as one address has been shared.

Unfortunately, with more than 100,000 dyadic exchanges, it was not feasible to determine definitively whether users kept in touch outside the dating site. Even the exchange of outside email addresses or phone numbers does not guarantee that the users actually called or wrote each other. Nonetheless, examining their messages for indications that they have provided this contact information gives at least a rough idea of how often users might have continued conversing.

Of the 110,723 dyadic conversations (sequence of one or more messages between distinct pairs of users) among users of the Site, 7.1 percent (7,879) appeared to contain a phone number — that is, they contained a string of the form "###-####," where # is any digit. Nearly half of the conversations (47.5 percent, or 52,606) appeared to contain an email address, indicated by a string that matched the form "x@x.x," where x is a series of any non-space, non-period characters.

Transition to Email

Of the exchanges containing an email address, 50.4 percent (26,508) consisted of just one unreciprocated message that included an email address. Men were more likely than women to send these unreciprocated messages containing an email address; they sent 72.2 percent of them (19,130).

Of the exchanges longer than one message that included an email address, an address was provided in the final message in the exchange 63.0 percent of the time. The man provided an address 69.3 percent of the time, while the woman provided an address less often, 48.8 percent of the time ($t = 38.026$, d.f. $= 25,724$, $p < 0.001$). Both the man and the woman provided an email address 19.1 percent of the time.

Transition to phone

Of the exchanges containing a phone number, 30.0 percent (2,367) consisted of just one unreciprocated message that included a phone number. As with email, men were more likely to send unreciprocated messages with a phone number; they sent 81.2 percent of them ($t = 38.928$; d.f. $= 2,366$; $p < 0.001$).

Of the reciprocated exchanges that included a phone number, a number was provided in the final message in the exchange 37.4 percent of the time. The man provided a phone number 74.5 percent of the time, whereas the woman provided a number 42.2 percent of the time (t = 23.730; d.f. = 3,827; p < 0.001). Both the man and the woman provided a phone number 17.1 percent of the time.

Impact on validity of *messages received*

The transition of dyads' communication to other channels makes *private messages received* a less than perfect metric of social attraction. As I have shown, this practice is prevalent among users of the Site. However, the presence of these transitions does not entirely corrupt *messages received* as a proxy for social attraction. Although it might be true that many successful dyads transition to phone or email, it is also true that in a dyadic interaction, the number of messages between the two users correlates positively with the inclusion of an email address and for the inclusion of a phone number (email: r = 0.145, p < 0.001; phone: r = 0.270, p < 0.001). If the provision of an address or phone number were prematurely terminating private messaging exchanges, we would expect an inverse, not direct, relationship between these variables.

Thus, I consider *messages exchanged* between a pair of users an imperfect but reasonable proxy for social affinity within a dyad. Generalizing beyond the dyad, the total *messages received* from all others serves as a proxy for a user's social attractiveness to aggregated other users from the Site.

Users continued to join the Site throughout the study period. To control for the duration of membership, we normalized *messages received* by *duration of membership* to realize our dependent metric for general social attractiveness, *messages received per day*.

Although we cannot infer strict causality in a post hoc analysis such as this, we can presume that, across nearly 60,000 active users, statistical interactions between self-descriptors in the profile and the number of messages the user receives result from differing responses to different values of a given self-descriptor or combination of self-descriptors. This is not experimentally verified, so the conclusions must be treated with caution, but logically the number of messages a user receives must depend on the actions of others who read the user's profile and choose to respond or not.

PREDICTING MESSAGES RECEIVED PER DAY

In this section, I will describe the relationship between the number of messages a user receives in the aggregate and his or her individual characteristics. To understand this relationship, I built linear regression models to predict the average *messages received per day* for men and women separately, following Marmaros and Sacerdote (2003).

Regression Model

I constructed the linear regression models in SPSS. The continuous variables *average messages sent per day* and *age* were used as they were; the other variables were categorical and so were converted into contrasts or ordered scales. *Educational level* ranges from 1 to 5, with the levels indicating, in order, high school, some college, college, some

Predictors of *messages received* for men in order of strength

Characteristic	Beta	
Avg. msgs sent per day	0.393	***
Age	0.147	***
Educational level	0.066	***
Physical attractiveness	0.049	***
Is Caucasian	0.046	***
Smokes	−0.044	***
Heavy build	−0.040	***
Is Christian	0.036	***
Has children	0.029	**
Never married	0.025	*
Has photo	0.023	**
Drinks	0.018	*

Predictors of *messages received* for women in order of strength

Characteristic	Beta	
Avg. msgs sent per day	0.404	***
Heavy build	−0.096	***
Physical attractiveness	0.078	***
Has photo	0.074	***
Age	0.063	***
Is Caucasian	0.060	***
Educational level	0.057	***
Smokes	−0.034	***
Drinks	0.032	***
Has children	0.026	*
Never married	0.026	*
Small build	0.020	*

$* p < 0.05$ $** p < 0.01$ $*** p < 0.001$

Table 2.3. Significant predictors of *messages received per day* from the regression models for male and female active users of the Site. *Beta* is the standardized regression coefficient for each characteristic. Not all characteristics had a significant effect in both models.

graduate work, and a completed graduate degree. *Physical attractiveness* ranges from 1 to 3, average to attractive to very attractive. The rest of the predictors are binary contrasts (−1 or 1). Contrasts for being Caucasian and being Christian were used because the majority of users on this Site were Caucasian and Christian; effectively, they indicate "majority" or "not majority."

After applying a log-transformation to the independent variable *messages received per day*, the residuals were normally distributed. Collinearity was low, with no tolerance below 0.500 in either the male or female model.

The full models are given in Appendix B, Table B.10, and residual diagnostics are in charts B.11 and B.12. Each model explains approximately 21 percent of the variance of *messages received per day* (R-squared = 0.21).

Strength of Predictors

Unsurprisingly, the most powerful predictor by far of how many messages a user receives is how many he or she sends. The next strongest predictor has less than half the effect of *messages sent per day*. Without this term in the models, they explain less than 5 percent of the variance in *messages received per day*. Clearly, how much effort users

expend in contacting others is a major determinant of how much attention they receive in return. It is in the less powerful predictors, however, that interesting but not unexpected gender differences emerge. Table 2.3 lists the significant predictors (p < 0.05) of *messages received* for men and women in descending order of strength.

Evolutionary Psychology

The field of evolutionary psychology has proposed a series of hypotheses to explain various aspects of human attraction. Typical theories suggest that women are attracted to men with the resources to care for them and their children, and men are attracted to women who appear capable of bearing healthy children (Buss 2003). Thus, for men, having a high income is attractive, while it is important for women to appear young and healthy, which correlates with fertility (Brehm et al. 2002).

Buss and Barnes (1986) conducted a series of interviews of married couples and found that the sexes preferred somewhat different sets of spousal characteristics. Those showing significant sex differences were *physically attractive* (valued more by men than by women), *college graduate*, and *good earning capacity* (both valued more by women than men). For women, physical attractiveness, including facial symmetry, correlates with health and thus fertility (Brehm et al. 2002). For men (or for anyone), more education is naturally associated with greater income.

Men

In this analysis, *age* and *educational level* had the greatest influence on the number of messages men received. Although income was not included in the data set, these two descriptors might serve as a proxy for it; the descriptor *educational level* is very similar to *college graduate* in Buss and Barnes (1986).

Whether a man was Christian had a moderate but highly significant effect on how many message he received, but, interestingly, the same effect was not present for women.

Wants children, *small build*, *athletic build*, and *height* were not significant predictors of *messages received* for men.

Women

The strongest predictors of how many messages women received were related to appearance — a heavy build was associated with receiving fewer messages, and physical attractiveness and posting a photograph were associated with receiving more. Interestingly, despite the negative influence of *heavy build*, a *small build* ("petite" or "slender") was not associated with appreciably more messages.

This result agrees with the findings of Buss and Barnes (1986) and other research in attraction. Men in industrialized societies have been found to prefer women of normal weight with a waist-to-hip ratio of approximately 0.7 (Marlowe and Wetsman 2001) — too heavy and too light are both less attractive. Also, men have been shown to be more concerned than women with the facial beauty of a potential partner (Brehm et al. 2002).

Of all the characteristics, *has photo* was ranked most differently for women and men. It was the fourth strongest predictor for women, but only the 11th strongest predictor — second to last — for men.

Wants children, is Christian, athletic build, and *height* were not significant predictors of *messages received* for women.

Commonalities

For both sexes, smoking reduced the expected number of messages received. Drinking increased it very slightly, more for women than for men; a large majority of users on the Site identified themselves as social drinkers.

In general, physical build did not greatly affect *messages received* unless the user in question chose one of the heavier build options ("a few extra pounds" or "full-figured"); this reduced the expected number of messages received, with a more serious impact on women than men. Both sexes valued self-reported physical attractiveness. Because it is self-reported, this descriptor is somewhat problematic. It is likely confounded with self-esteem and modesty.

Life-course characteristics, such as being previously married or having children, had only small effects on the expected number of messages received. *Wants children* was not significant in either the male or the female model. At first, this is surprising, because life-course attributes were found to be the most strongly bounding characteristics in the earlier analysis of dyadic contacts — users clearly sought people like them on these dimensions.

In fact, it is because they are strongly bounding that no particular life-course choice has much influence on the aggregate number of messages received; people who want children are adamant about finding others who share their preference, but people who do not want children are equally steadfast in seeking a partner with the same idea. When "child" people contact other "child" people and "no-child" people contact other "no-child" people, child preferences will not have a strong effect in a regression model derived in this manner.

Because they derive from the aggregated behavior of thousands of users, the most influential characteristics in these regression models are those on which many users felt similarly — for example, that physical attractiveness was desirable.

Explained Variation in *Messages Received per Day*

The total explained variation in these models, especially without the *messages sent* term, might seem low. Although the regression models could probably be improved slightly by a skilled statistician, I believe the explained variation is low for two reasons.

First, users do not (and cannot) look at all the profiles on the Site before deciding whom to contact. Thus, although the people they select might be optimal within the subset of profiles they have examined, they will not be globally optimal across the set of all users — that is, even if a user selects the most intelligent person in the subset they consider, it is likely that someone in the full set of profiles has even greater intelligence. By definition, the small subset that any user can consider is therefore inadequate for them to express their preferences fully through their selection. Thus, it is uncertain how much variance the models could possibly explain, but it is clearly above what is stated in the current model (which assumes a consideration of the whole population) and less than 100 percent.

Second, the characteristics that we are able to analyze quantitatively might not capture a great deal of the information that users employ to make decisions about whom to contact. Human assessments of the free-text

components and photos in profiles would be necessary to build a more complete model — and even then, attraction can be idiosyncratic.

Furthermore, attraction is is likely to be a variable with high heterogeneity. Divining a correct, universally applicable model of how much attention others will pay to a particular person is not realistic.

Chapter 3

Future Directions for Online Dating

DESIGN IMPLICATIONS

The behavioral analysis in the previous chapter can aide efforts to improve online dating systems in the future. First, we can use findings about which characteristics users rely on to make decisions about whom to contact to make simpler profiles that comprise only the truly salient information. Some descriptors are noise more than signal: they clutter the profile with information that users do not need, and they attract attention that might otherwise be spent on more salient attributes.

Moreover, to the degree that dating sites use all of their profile information, even the extraneous, to perform automated matching, we can make the matching process more accurate by weighting more heavily those features that users deem relevant through their actions in studies like this.

Specifically, it appears that descriptors relating to life course are among the most important pieces of information to learn about a prospective partner. Online dating systems should bring this information to the forefront in their designs; they should also give it additional weight in their matching algorithms.

This study also highlighted dramatic differences in preference by gender. Such differences are not surprising, but present heterosexual dating systems do not tailor their searching and matching tools to the distinct needs of men and women. Table 2.3 indicates how much the most desirable attributes differ by sex; search tools should present interfaces that reflect the varying goals of the groups they serve.

It is likely that other groups, perhaps divided according to culture, religion, or geography, have similarly distinct and specific needs. The data set in this study was unsuitable for answering that question, however, because it was mostly Caucasian, Christian, and American. Future research should examine this issue and provide guidance for designers of dating systems.

The inability of this analysis to explain a large proportion of the variance in who contacts whom suggests that simple categorical and numerical characteristics provide only part of the information that people use to seek dates online. Researchers will need other methods — ethnography, surveys — to determine what else influences people in their search and communication. In the New Directions section below, I offer some ideas about how we might move our designs beyond the database mentality as well.

OTHER DESIGN CONSIDERATIONS

The Dilemma of Success

Two conflicting goals face designers of online personals systems:

1. Support users who want to meet a partner with whom they can have a successful relationship.

2. Support the growth of the site by retaining users for as long as possible.

These goals conflict because a user who succeeds in finding a good match will no longer need the site. A system

that proves too successful in this sense might not receive the revenue it needs to stay in business. Even a non-commercial system needs to retain a critical mass of users to remain useful. In considering the design of personals systems, we should keep in mind the tension between these goals. We might discover that we can serve one only at the expense of the other.

Of course, this dilemma remains speculation until designers are able to create a site that pairs people successfully enough to realize it. Even then, a population that continually adds as well as loses users could likely sustain the business of a site. The conflicting goals become problematic only if designers prioritize the business model over user satisfaction, creating a system that facilitates, e.g., serial short-term relationships — and thus heavy usage of the site — while leading users to believe that it will help them find an ongoing long-term relationship.

Designing for Distinct Populations

Thompson (2003) claims that designers of "social software" focus too much on building generalized systems for communication, exhibiting a "desire for simplification and regimentation instead of seeking to understand complexity." In designing online dating systems, we must consider whether a single design can work for people variously seeking marriage, casual dates, casual sex, and friends. Perhaps a more articulated system for each purpose would better serve each population. It is likely that specific needs vary according to other groupings as well: ethnicity, religion, sexual orientation.

Already, some dating sites have specialized, adding population-specific details to their profiles and search options. The first site for a specific subpopulation was JDate, for Jewish users, which was founded by the rabbi who also created speed dating. In addition to providing a large pool of Jewish singles (who might be a small minority on many sites), JDate targets its population by allowing users to specify their sect of Judaism in profiles and searches. Similarly, Manhunt.net, a site for gay men, offers specialized physical build options such as "bear," a body description unique to gay male culture. Additionally, some unspecialized sites also run smaller sub-sites for subpopulations.

Most subpopulation sites limit their specialization to the addition of descriptors specific to their target users. In the future, designers might also tailor interactional tools to specific audiences, should it be established that some tools serve some audiences better, as seems likely.

NEW DIRECTIONS FOR ONLINE DATING

The future of online dating will be shaped both by new technologies and by novel designs. Humans have a wealth of faculties for comprehending and communicating with others, but present designs and technologies take advantage of only a subset of them. This section describes some of the ways online dating could grow in the future.

Just-in-Time and Just-in-Place

Mobile technologies for identifying people nearby who are looking for someone like you have already emerged. A Japanese gadget called Lovegety, popular in the late 1990s, used short-range wireless communication to compare

profile and sound an alert when a compatible other with the same device is near. These devices never achieved a large enough user population to realize their potential; without a sufficient density of people with the devices, a user might never come in contact with another user (Iwatani 1998).

More recent efforts, like the Media Lab's Serendipity project (Eagle et al. 2004), attempt to fuse the social network concepts behind tools like Friendster with the just-in-time, just-in-place aspect of the Japanese devices. Serendipity uses location-aware mobile phones to identify friends of friends who might like to meet. Because it tracks user location with a central server, it avoids the problem of short-range, peer-to-peer devices that might never come in close enough proximity with another device. However, as it is deployed, it will still face the same critical mass problem than online dating Web sites have already overcome; with Serendipity, one can identify only other Serendipity users. The success of systems like Serendipity, some of which are also being developed by online dating companies, is unclear as of this writing.

Experiential Online Dating

The focus of online dating systems on explicit self-description and deliberate dating-oriented interactions feels forced to some users. New designs for dating systems could facilitate interaction through mutual exploration, play, or creation — collaborative story-writing, puzzles, virtual environments to wander together.

Users might gain more authentic information about each other in a space in which gaining such information is not the predominant task. People are notoriously bad at describing themselves accurately, so a good design would give them a chance to show who they are rather than having to explain it. Furthermore, putting the focus on an external object or task more closely mimics real-world interaction, reducing the artificiality that some find makes online personals an awkward medium.

The Biology of Chemistry

Subconsciously processed, often unintentional produced signals play a major role in face-to-face attraction: body language, fleeting facial expressions, intonation, scent, pheromones. These kinds of cues tend to be given off inadvertently, in Goffman's (1959) terminology, rather than deliberately given. Some of them, like microexpressions (Ekman & Friesen 1969) and pheromonal signals, are impossible to control consciously. They have no analogue in textual online environments, where users have the time to craft their message via a medium that reveals no more than they intend to show.

Pheromones play a poorly understood role in human attraction. Mammals detect pheromones with sense organs near to but distinct from those in the nose used for smelling. Laboratory experiments demonstrate that they affect how attractive people find one another, but the strength of this effect compared to other determinants of attraction is unclear. In experiments in which women were exposed to men's axillary sweat but never met the men, women preferred the scent of men whose genotypes were similar to her father's genotype, but not so similar as to be incestuous (Jacob et al. 2002). The evolutionary explanation for this tendency is that women would prefer to mate with someone whose genotype is as much like their own as possible without being so close as to lead to the genetic defects that come from inbreeding. Similarly, women preferred men with whom they would be likely to produce

genetically healthy children (Wedekind et al. 1995).

This kind of information cannot be captured or conveyed by current technology and so cannot be shared with potential partners in an online dating environment. Online dating users cannot perceive pheromonal attraction until they meet in person, by which time they might have made a significant investment in terms of time or emotional energy with the other person. Although it is unclear exactly what the matching algorithm would be to discover pheromonally compatible pairs of people, it seems likely given the basis in genetics that at least a crude algorithm could be discovered given more empirical work. Technology will emerge to digitize biological signals like pheromones, enabling widespread collection of this information, which could be used for searching and matching. However, because it is uncertain how great a role pheromonal affinity plays in human attraction, we cannot be sure how much such information would improve the matching capabilities of online dating systems.

NEW DIRECTIONS FOR RESEARCH IN MEDIATED DATING

The present study represents one approach to one data set, analyzed *post hoc*. Out of curiosity as social scientists and desire for guidance as designers, we can imagine a variety of future research that would further our understanding of how people use online dating systems.

Ethnographic Study

Online dating is not an isolated act; users integrate these systems into their lives for various purposes and lengths of time. Examining only the online portion of a user's behavior provides an incomplete picture of the total experience.

Longitudinal study of people who use online personals will answer questions about the connection between relationships fostered through the Internet and the user's offline social milieu. In this area, no existing work moves beyond anecdote. We do not know how users think about and relate to the people they meet via online personals — does the fading cultural stigma affect perceptions of their dates? We do not know how satisfied people are with their dates, whether they live up to expectations. We do not know how the process of searching and meeting in person repeatedly shapes the user's approach to using dating systems. Questions like these require in-depth, long-term study of online dating users.

Comparative Analysis

The populations of online dating systems are as diverse as the sites themselves. The users of the Site examined in this work consist of only Americans and Canadians. Most of the active users appear to reside in rural, suburban, and secondary urban areas; major urban areas like New York and Los Angeles appear to be under-represented. But the exact nature of the differences is not important — the result is that the users and activity on this Site, or any one site, do not necessarily represent those of users elsewhere. We cannot assume that the specific results in the present study, even though they draw from a large set of users and interactions and thus have great statistical power, will hold for other sites and other populations.

An important part of the research agenda should be the comparison of different online dating sites — their

member composition, the tools they provide, and the way users employ those tools. It would be especially valuable to compare the true efficacies of various sites. Of course, standards for efficacy vary with the goal of the site and the users: for some, it might be fostering many successful long-term relationships, but for others, it might be providing a large number of attractive sex partners.

It be particularly useful to compare the communication tools provided by different sites. Most private messaging systems are just light-weight email systems, but some, like eHarmony's, provide other mechanisms for interaction. Researchers could compare these tools with more traditional approaches to determine whether they facilitate the kind of communication users need to make decisions about whom they want to meet in person.

At the same time, it is important to remember that the medium is not the primary determinant of what people will do in computer-mediated communication. To borrow Walther et al.'s (2001) observation about their results: "[Q]ualities of CMC are, in this case as in others, more often the product of interesting and predictable interactions of several mutual influences than main effects of media."

Experimental Work

Norton, Frost, and Ariely (2004) have begun to conduct experiments with online dating scenarios to determine how users perceive others in these environments. Studies like theirs are important for establishing rigorously the psychology at work when users evaluate and communicate with potential partners online. *Post hoc* analyses such as the present work, and indeed any natualistic observation, have the advantage of context, but they lack the rigor of laboratory experiments with controlled conditions.

Frost, Ariely, Donath, and I have proposed a study to examine the effects of photographs. As Walther et al. (2001) show, photographs play a powerful role in establishing social presence. We want to explore how different photos of the same person with the same profile affect how others perceive them. To accomplish this, we propose shooting photos of the same subject with different settings and styles, and then posting online personal ads with each of the photos. We would compare the quantity and variety of messages each ad attracts. Additionally, we would bring subjects into the laboratory to evaluate the photos in terms of attractiveness and message.

Analytic Work

Finally, the present work can be extended and supplemented by further analysis of the Site I have studied and by similar analysis of other systems. Few data sets provide such a wealth of information about human romantic interaction. With such a large number of cases, too, statistical analyses are extraordinarily sensitive even to small effects. This requires us to consider strength as well as significance, but with a hypothesis in mind, subtle effects can be detected. The disadvantage of large-scale quantitative analysis alone, of course, is that it misses the experiential facet of online dating, which is needed to complete our understanding of the process.

One topic not addressed here is searching behavior; I have discussed what characteristics actually seek, but we do not know whether their searching processes support finding what users appear to want. Also, users' behavior before they quit the dating site — either explicitly, or implicitly by not showing up for a long time — might yield insight

into whether they are leaving because they found a partner or because they are dissatisfied or bored.

Some degree of manual coding of profile features and messaging exchanges would improve quantitative analyses, although it would limit the scale of the sample. Particularly useful features to code would be attractiveness of photographs and style of writing in free-text sections of the profile.

IN CONCLUSION

My exploration of behavior in online dating environments raises more questions than it answers. Each statistic begs for comparison with other online dating systems, with different populations and different interactional tools. The numerical results have human stories behind them that will require a complementary qualitative approach to unravel.

Of course, the real question is what people are looking for in a partner, whether for a night, a date, or a lifetime. Online dating systems provide a copious amount of raw behavioral data that might improve our understanding of people's complex goals in this arena, which will in turn help us improve the systems themselves.

Many tools for computer-mediated communication seem to have been designed more for the convenience of the engineer than for the human needs of the user; they tend to lose the subtleties of social perception and communication in favor of the structure of a relational database. As I examined the data for this work, I began to wonder what really matters in attraction and love — and can a set of simple descriptors capture even a sliver of it? The analysis suggests that they can indeed capture a sliver, but the whole pie may yet elude them. Even the characteristics that were found to be important must be considered with some skepticism: just because they were found to be the most important characteristics *of those provided by the Site* does not mean that many more important factors do not exist.

The task of those who will research and design online dating systems in the future is to help people find what they really want, or at least what they think they want — the people who succeed will gain the satisfaction of making a lot of people just a little bit happier.

References

Ahuvia, A.C., & M.B. Adelman. "Formal Intermediaries in the Marriage Market: A Typology and Review." In J. of Marriage and the Family 54 (May 1992): 452-463.

Ariely, D. Personal correspondence. 2003.

Armstrong, J. Conditions of Love: The Philosophy of Intimacy. New York: Penguin Books, 2002.

Batty, D. and J. McCurry. "Children to be shielded from abuse via mobiles." *The Guardian*, Jan. 12, 2004. Available: http://www.guardian.co.uk/online/story/0,3605,1121078,00.html

Becker, G. A Treatise on the Family. Cambridge, Mass.: Harvard University Press, 1992.

Bolig, R., Stein, P.J., & P.C. McKenry. "The Self-Advertisement Approach to Dating: Male-Female Differences." In Family Relations 33 (1984): 587-592.

Botti, S. & S.S. Iyengar. "The psychological pleasure and pain of choosing: When people prefer choosing at the cost of subsequent satisfaction." Under review for *Journal of Personality and Social Psychology*, 2003.

Brehm, Miller, Perlman, and Campbell (2002). Intimate Relationships 3e. New York: McGraw-Hill Higher Education.

Buss, D. *The Evolution of Desire: Strategies of Human Mating (Revised Edition 4)*. New York: Basic Books, 2003.

Buss, D. and M. Barnes. "Preferences in human mate selection." In *Journal of Personality and Social Psychology*, 50 (1986): 559-570.

Clement, R. W., & Krueger, J. (2002). "Social categorization moderates social projection." In *Journal of Experimental Social Psychology*, 38, pp. 219-231.

Donath, J. "Identity and Deception in the Virtual Community." In *Communities in Cyberspace*, P. Kollock and M. Smith, eds. London: Routledge, 1998.

Duda, R.O., Hart, P.E., and D.G. Stork. *Pattern Classification, 2e*. New York: John Wiley & Sons, Inc., 2001.

Eagle, N. Serendipity project. 2004. Available: http://www.mobule.net/

Edmonds, B. "Capturing Social Embeddedness: A Constructivist Approach." October 1998. Available: http://bruce.edmonds.name/socemb/

Ekman, P., & Friesen, W V. (1969). Nonverbal leakage and clues to deception. Psychiatry. 32. 88-105.

Farnham, S. & J. Riegelsberger. "Social Engineers: Enabling the Matchmaking Game in Online Dating Systems." Position paper, Computer-Human Interaction 2004 workshop, "Scientists, designers seek same for good conversation': A Workshop on Online Dating."

Fiore, A., Frost, J. & J. Donath. "'Scientists, designers seek same for good conversation': A Workshop on Online Dating." In *Computer-Human Interaction 2004 Extended Abstracts*. 26 April 2004, Vienna, Austria.

Fisher, H. Anatomy of Love: A natural history of mating, marriage, and why we stray. New York: Fawcett-Columbine, 1992.

Frank, R. "Chapter 10: Love." *Passions Within Reason*. New York: WW Norton, 1988.

Gay, G. and H. Hembrooke. *Activity-Centered Design: An Ecological Approach to Designing Smart Tools and Usable Systems*. Cambridge, Mass.: MIT Press, 2004.

Gilding, M. "Families of the New Millennium." In *Family Matters*, No. 62, Winter 2002.

Goffman, Erving. *The Presentation of Self in Everyday Life*. New York: Doubleday, 1959.

Goodwin, R. "Dating agency members: Are they 'different'?" In *J. of Social and Personal Relationships*, 7 (1990): 423-430.

Grigoriadis, V. "The New Position on Casual Sex." In *New York Metro*, Jan. 13, 2003.

Holme, P., Edling, C.R., & F. Liljeros. "Structure and Time-Evolution of an Internet Dating Community." v. 2, 15 October 2003.

Iwatani, Y. "Love: Japanese Style." *Wired News*, Jun. 11, 1998. Available: http://www.wired.com/news/culture/0,1284,12899,00.html

Iyengar, S. and M. Lepper. "When choice is demotivating: Can one desire too much of a good thing?" In *Journal of Personality and Social Psychology*, 76 (2000): 995-1006.

Jacob, S., McClintock, M. K., Zelano, B. and Ober, C. "Paternally inherited HLA alleles are associated with women's choice of male odor." In Nature Genetics, 30 (2002): 175-179.

Lea, M., & Spears, R. Love at first byte? Building personal relationships over computer networks. In J. T. Wood & S. Duck (Eds.), Under-studied relationships: Off the beaten track (pp. 197-233). Thousand Oaks, CA: Sage, 1995.

Marlowe, F. and A. Wetsman. "Preferred waist-to-hip ratio and ecology." *Personality and Individual Differences*, 30 (2001): 481-489.

Marmaros, D. and B. Sacerdote. "How Friendships Form." Sept. 15, 2003.

McKenna, K.Y.A., Green, A.S., & M.E.J. Gleason. "Relationship Formation on the Internet: What's the Big Attraction?" In *J. of Social Issues*, Vol. 58, No. 1, 2002, pp. 9-31.

Mulrine, A. "Love.com: For better or for worse, the Internet is radically changing the dating scene in America." U.S. News & World Report, Sept. 29, 2003.

Neils, J. "The Internet as a Medium for Meeting: An Analysis of the Cultural Effects of Online Dating." July 25, 2003. Available: http://students.washington.edu/jneils/onlinedating/

Norton, M., Frost, J. & D. Ariely. "Virtue in Vagueness: Ambiguity and Online Dating." Position paper, Computer-Human Interaction 2004 workshop, "'Scientists, designers seek same for good conversation': A Workshop on Online Dating."

Nussbaum, E. "Net Gain: A Pollyanna-ish view of online personals." In *Slate.com*, June 18, 2002. Available: http://slate.msn.com/?id=2067095

Ott, R.L. and M. Longnecker. *An Introduction to Statistical Methods and Data Analysis, 5e.* Pacific Grove, Calif.: Duxbury, 2001.

Priyanka, V. "Indians Turn To Internet For Spouses." *Media General News Service,* June 7, 2004.

Sears, David O. 1983. "The Person-Positivity Bias." In *Journal of Personality and Social Psychology,* 44: pp. 233-250.

Shibazaki, K. & K.A. Brennan (1998). "When birds of different feathers flock together: A preliminary comparison of intraethnic and inter-ethnic dating relationships." J. of Social and Personal Relationships, 15, 248-256.

Sydney Morning Herald. "Japanese police to regulate Internet dating services." Dec. 26, 2002. Available: http://www.smh.com.au/articles/2002/12/26/1040511126218.html

Taylor, E. & L. Sharkey. "A Very Personal Make-Over." *Nerve.com,* Sept. 27, 2001. Available: http://personals.nerve.com/Content.aspx?feed=emlo/&article=20010927

Taylor, E. & L. Sharkey. "A Very Personal Make-Over: The Sequel" *Nerve.com,* Jan. 23, 2003. Available: http://personals.nerve.com/Content.aspx?feed=emlo/&article=20030123

Thompson, B. "A little less conversation." *BBC News,* 30 May 2003. Available: http://news.bbc.co.uk/1/hi/technology/2949588.stm

U.S. Census Bureau. "Number, Timing and Duration of Marriages and Divorces: 1996." February 2002. Available: http://www.census.gov/prod/2002pubs/p70-80.pdf

U.S. Census Bureau. "Number, Timing, and Duration of Marriages and Divorces in the United States: June 1975." October 1975. Available: http://www.census.gov/population/www/socdemo/marr-div/p20-297.html

United Nations Department of Economic and Social Affairs. *Demographic Yearbook: 1997.* New York: United Nations, 1999.

Vaisman, D. "*Nerve* Personals: The White Hipster's Last Romantic Refuge." *New York Press*, Vol. 14, No. 3, 2001. Available: http://www.nypress.com/14/3/news&columns/nyc2.cfm

Van Gelder, L. "The strange case of the electronic lover." *Ms.*, October 1985, pp. 94-104, 117, 123-124.

Walther, J. B. "Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction." In *Comm. Research* 23, 3-43, 1996.

Walther, J.B., Slovacek, C.L., & L.C. Tidwell. "Is a Picture Worth a Thousand Words? Photographic Images in Long-Term and Short-Term Computer-Mediated Communication." In *Comm. Research* 28 No. 1, February 2001, pp. 105-134.

Wedekind, C., Penn, D. "MHC genes, body odours, and odour preferences." In *Nephrol Dial Transplant* (2000) 15: 1269-1271.

Wedekind, C., Seebeck, T., Bettens, F., and A.J. Paepke. "MHC-Dependent Mate Preferences in Humans." In *Proc Biological Sci.*, vol. 260, no. 1359 (June 1995): pp. 245-49

Wellman, B., & M. Gulia. "Net Surfers Don't Ride Alone: Virtual Community as Community." In B. Wellman (Ed.) Networks in the Global Village (pp. 331-67). Boulder, Colo.: Westview Press, 1999.

Williams, L.M. & M.G. Lawler. "Marital Satisfaction and Religious Heterogamy: A Comparison of Interchurch and Same-Church Individuals." In *J. of Family Issues* 24, No. 8, 2003: 1070-1092.

# Appendix A

Demographic Statistics

A.1 ACTIVE USERS BY STATE OR PROVINCE

Rank	State	Users	Percent	Cumul. Percent	U.S. State Pop. Rank[1]
1	New York	10348	18.1	18.1	3
2	California	2990	5.2	23.3	1
3	Wisconsin	2979	5.2	28.5	18
4	North Carolina	2858	5.0	33.5	11
5	Illinois	2506	4.4	37.8	5
6	Nevada	2448	4.3	42.1	35
7	Maine	2443	4.3	46.4	40
8	Iowa	2054	3.6	50.0	30
9	Georgia	1942	3.4	53.3	10
10	Pennsylvania	1836	3.2	56.5	6
11	South Carolina	1671	2.9	59.5	26
12	Florida	1578	2.8	62.2	4
13	Saskatchewan (Can.)	1558	2.7	64.9	
14	Arizona	1531	2.7	67.6	20
15	Oregon	1322	2.3	69.9	28
16	Louisiana	1245	2.2	72.1	22
17	Minnesota	1114	1.9	74.0	21
18	Missouri	1068	1.9	75.9	17
19	Ohio	985	1.7	77.6	7
20	Indiana	925	1.6	79.2	14
21	Texas	911	1.6	80.8	2
22	Nebraska	805	1.4	82.2	38
23	Tennessee	759	1.3	83.5	16
24	Alabama	730	1.3	84.8	23
25	Kentucky	714	1.2	86.1	25
26	New Jersey	713	1.2	87.3	9
27	Oklahoma	698	1.2	88.5	27
28	Washington	689	1.2	89.7	15
29	Michigan	640	1.1	90.8	8
30	New Mexico	602	1.1	91.9	36
31	Alaska	494	.9	92.8	48
32	Massachusetts	442	.8	93.5	13
33	Mississippi	393	.7	94.2	31
34	New Hampshire	372	.6	94.9	41
35	Vermont	305	.5	95.4	49
36	Hawai'i	287	.5	95.9	42
37	Virginia	267	.5	96.4	12
38	Ontario (Can.)	217	.4	96.7	
39	Colorado	207	.4	97.1	24
40	West Virginia	193	.3	97.4	37
41	Arkansas	192	.3	97.8	33
42	Connecticut	184	.3	98.1	29
43	Idaho	178	.3	98.4	39
44	Utah	161	.3	98.7	34
45	Maryland	142	.2	98.9	19
46	Kansas	141	.2	99.2	32

47	Alberta (Can.)	83	.1	99.3	
48	Rhode Island	69	.1	99.4	43
49	Montana	46	.1	99.5	44
50	British Columbia (Can.)	41	.1	99.6	
51	Nova Scotia (Can.)	32	.1	99.7	
52	South Dakota	31	.1	99.7	46
53	Newfoundland (Can.)	23	.0	99.7	
54	North Dakota	17	.0	99.8	47
55	Delaware	16	.0	99.8	45
56	District of Columbia	15	.0	99.9	50
57	Armed Forces Africa	13	.0	99.9	
58	Puerto Rico	13	.0	99.9	
59	Quebec (Can.)	13	.0	99.9	
60	Manitoba (Can.)	12	.0	99.9	
61	Nebraska	10	.0	100.0	38
62	Wyoming	9	.0	100.0	51
63	Armed Forces Pacific	2	.0	100.0	
64	Other	24	.0	100.0	
	Total	**57307**	**100.0**		

1. From Demographica (2000) http://www.demographia.com/db-2000stater.htm

A.2 NUMBER OF ACTIVE USERS BY AGE AND GENDER
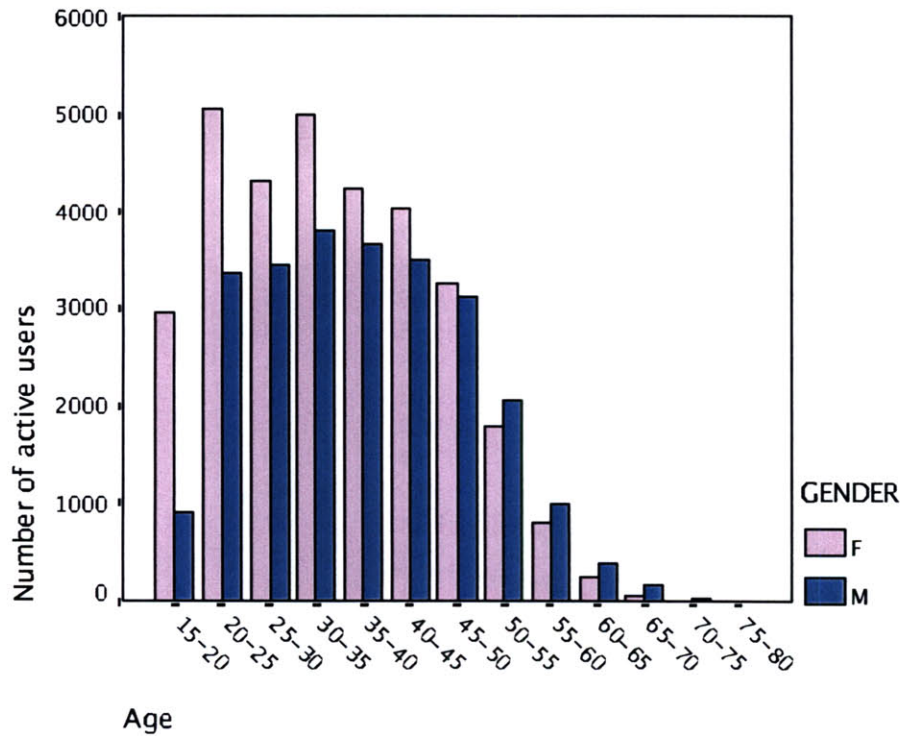
A.3 POSSIBLE VALUES FOR CATEGORICAL CHARACTERISTICS

Build

Petite

Slender

Average

Athletic

Few extra pounds

Full figured

Proportional

Body builder

Tall and lanky

No preference/Prefer not to answer

Drinking Habits

Doesn't drink

Occasionally/Socially

Regularly

Trying to quit

No preference/Prefer not to answer

Education Level

High school

Some college

College graduate

Some post-college

Masters

Doctorate

Graduate Degree

No preference/Prefer not to answer

Has Children

No children

1

2

3

4

5+

No preference/Prefer not to answer

Marital Status

Never married

In a relationship

Married

Divorced

Separated

Widowed

No preference/Prefer not to answer

Pets

Has pet(s)

No pet(s), but likes pets

No pet(s); can't be around pets

No preference/Prefer not to answer

Pets Owned

A dog owner

A cat owner

A reptiles owner

A fish owner

A rodent owner

A bird owner

Call me Old McDonald

Petless

No preference/Prefer not to answer

Physical Appearance

Very attractive
Attractive
Average
No preference/Prefer not to answer

Race

Caucasian
African Amercian
Asian
Multi-racial
Hispanic
East Indian
Other
No preference/Prefer not to answer

Relationship Sought

E-mail
Friends
Casual/Short-term
Short-term or Long-term
Long-term relationship
Marriage
Possible marriage
Friends first
Travel partners
Seniors
Sports partners
Physical/Intimate
No preference/Prefer not to answer

Religion

Christian
Jewish
Muslim
Hindu
Agnostic
Atheist
Other
No preference/Prefer not to answer

Smoking habits

Doesn't smoke
Occasionally/Socially
Regularly
Trying to quit
No preference/Prefer not to answer

Wants Children

Does not want children
Wants children
Undecided
No preference/Prefer not to answer

A.4 USERS BY GENDER AND RACE

			Caucasian	African-American	Asian	Multi-racial	Hispanic	East Indian	Other	Prefer not to answer	Total
							Race				**Total**
Gender	F	Count	24032	653	259	318	657	20	423	2510	28872
		Exp. Count	24174.1	642.2	204.9	265.6	560.7	21.9	355.8	2646.8	28872.0
		% within Gender	83.2%	2.3%	.9%	1.1%	2.3%	.1%	1.5%	8.7%	100.0%
		% within Race	54.3%	55.6%	69.1%	65.4%	64.0%	50.0%	65.0%	51.8%	54.7%
		% of Total	45.5%	1.2%	.5%	.6%	1.2%	.0%	.8%	4.8%	54.7%
	M	Count	20201	522	116	168	369	20	228	2333	23957
		Exp. Count	20058.9	532.8	170.1	220.4	465.3	18.1	295.2	2196.2	23957.0
		% within Gender	84.3%	2.2%	.5%	.7%	1.5%	.1%	1.0%	9.7%	100.0%
		% within Race	45.7%	44.4%	30.9%	34.6%	36.0%	50.0%	35.0%	48.2%	45.3%
		% of Total	38.2%	1.0%	.2%	.3%	.7%	.0%	.4%	4.4%	45.3%
Total		Count	44233	1175	375	486	1026	40	651	4843	52829
		Exp. Count	44233.0	1175.0	375.0	486.0	1026.0	40.0	651.0	4843.0	52829.0
		% within Gender	83.7%	2.2%	.7%	.9%	1.9%	.1%	1.2%	9.2%	100.0%
		% within Race	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	83.7%	2.2%	.7%	.9%	1.9%	.1%	1.2%	9.2%	100.0%

A.5 RELIGION OF ACTIVE USERS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	(Invalid)	28	.0	.1	.1
	Christian	28996	50.6	54.9	55.0
	Jewish	487	.8	.9	55.9
	Muslim	32	.1	.1	56.0
	Hindu	31	.1	.1	56.0
	Agnostic	932	1.6	1.8	57.8
	Atheist	405	.7	.8	58.5
	Other	3826	6.7	7.2	65.8
	Prefer not to answer	18060	31.5	34.2	100.0
	Total	52797	92.1	100.0	
Missing	System	4510	7.9		
Total		57307	100.0		

A.6 MARITAL STATUS OF ACTIVE USERS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Never married	19903	34.7	37.7	37.7
	In a relationship	746	1.3	1.4	39.1
	Married	757	1.3	1.4	40.5
	Divorced	20828	36.3	39.4	79.9
	Separated	2948	5.1	5.6	85.5
	Widowed	1394	2.4	2.6	88.2
	Prefer not to answer	6254	10.9	11.8	100.0
	Total	52830	92.1	100.0	
Missing	System	4533	7.9		
Total		57363	100.0		

A.7 EDUCATIONAL LEVEL OF ACTIVE USERS

	Frequency	Percent	Cumulative Percent
High school	8095	15.3	15.3
Some college	19422	36.8	52.1
College graduate	12368	23.4	75.5
Some post-college	2280	4.3	79.8
Masters	47	.1	79.9
Doctorate	6	.0	79.9
Graduate degree	3794	7.2	87.1
Prefer not to answer	6811	12.9	100.0
Total	52823	100.0	100.0

A.8 DRINKING HABITS OF ACTIVE USERS

	Frequency	Percent	Cumulative Percent
Doesn't drink	6085	11.5	11.5
Occasionally/Socially	40608	76.9	88.4
Regularly	935	1.8	90.1
Trying to quit	79	.1	90.3
Prefer not to answer	5125	9.7	100.0
Total	52832	100.0	100.0

A.9 SMOKING HABITS OF ACTIVE USERS

	Frequency	Percent	Cumulative Percent
Doesn't smoke	31900	60.4	60.4
Occasionally/Socially	5822	11.0	71.4
Regularly	5603	10.6	82.0
Trying to quit	3838	7.3	89.3
Prefer not to answer	5663	10.7	100.0
Total	52826	100.0	100.0

A.10 NUMBER OF CHILDREN OF ACTIVE USERS

	Frequency	Percent	Cumulative Percent
No children	23731	44.9	44.9
1	8194	15.5	60.5
2	9180	17.4	77.8
3	3663	6.9	84.8
4	900	1.7	86.5
5+	289	.5	87.0
Prefer not to answer	6849	13.0	100.0
Total	52806	100.0	100.0

A.11 *WANTS CHILDREN* PREFERENCES OF ACTIVE USERS

	Frequency	Percent	Cumulative Percent
Does not want children	12549	23.8	23.8
Wants children	12914	24.5	48.2
Undecided	11511	21.8	70.0
Prefer not to answer	15832	30.0	100.0
Total	52806	100.0	100.0

A.12 ACTIVE USERS BY GENDER AND BUILD

		BUILD										Total
		Petite	Slender	Average	Athletic	Few extra pounds	Full figured	Proportional	Body builder	Tall and lanky	Prefer not to answer	
Fem.	Count	3228	3668	10393	2667	2883	2155	1299	33	81	2463	28870
	Expected Count	1796.8	2856.9	10589.8	5720.3	2977.6	1272.7	1019.1	149.7	153.0	2331.2	28870.0
	% within Gender	11.2%	12.7%	36.0%	9.2%	10.0%	7.5%	4.5%	.1%	.3%	8.5%	100.0%
	% within BUILD	98.2%	70.2%	53.6%	25.5%	52.9%	92.5%	69.7%	12.0%	28.9%	57.7%	54.6%
	% of Total	6.1%	6.9%	19.7%	5.0%	5.5%	4.1%	2.5%	.1%	.2%	4.7%	54.6%
Male	Count	60	1560	8986	7801	2566	174	566	241	199	1803	23956
	Expected Count	1491.2	2371.1	8789.2	4747.7	2471.4	1056.3	845.9	124.3	127.0	1934.8	23956.0
	% within Gender	.3%	6.5%	37.5%	32.6%	10.7%	.7%	2.4%	1.0%	.8%	7.5%	100.0%
	% within BUILD	1.8%	29.8%	46.4%	74.5%	47.1%	7.5%	30.3%	88.0%	71.1%	42.3%	45.4%
	% of Total	.1%	3.0%	17.0%	14.8%	4.9%	.3%	1.1%	.5%	.4%	3.4%	45.4%
Total	Count	3288	5228	19379	10468	5449	2329	1865	274	280	4266	52826
	Expected Count	3288.0	5228.0	19379.0	10468.0	5449.0	2329.0	1865.0	274.0	280.0	4266.0	52826.0
	% within Gender	6.2%	9.9%	36.7%	19.8%	10.3%	4.4%	3.5%	.5%	.5%	8.1%	100.0%
	% within BUILD	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
	% of Total	6.2%	9.9%	36.7%	19.8%	10.3%	4.4%	3.5%	.5%	.5%	8.1%	100.0%

A.13 USERS BY GENDER AND SELF-REPORTED PHYSICAL APPEARANCE

|		Physical appearance				Total
		Very attractive	Attractive	Average	Prefer not to answer	
Female	Count	3113	16181	6093	3478	28865
	Expected Count	2474.6	15649.3	7305.2	3435.9	28865.0
	% within Gender	10.8%	56.1%	21.1%	12.0%	100.0%
	% within Appear.	68.8%	56.5%	45.6%	55.3%	54.7%
	% of Total	5.9%	30.6%	11.5%	6.6%	54.7%
Male	Count	1415	12454	7274	2809	23952
	Expected Count	2053.4	12985.7	6061.8	2851.1	23952.0
	% within Gender	5.9%	52.0%	30.4%	11.7%	100.0%
	% within Appear.	31.3%	43.5%	54.4%	44.7%	45.3%
	% of Total	2.7%	23.6%	13.8%	5.3%	45.3%
Total	Count	4528	28635	13367	6287	52817
	Expected Count	4528.0	28635.0	13367.0	6287.0	52817.0
	% within Gender	8.6%	54.2%	25.3%	11.9%	100.0%
	% within Appear.	100.0%	100.0%	100.0%	100.0%	100.0%
	% of Total	8.6%	54.2%	25.3%	11.9%	100.0%

Appendix B

Interactional Statistics

B.1 MEAN NUMBER OF TIMES USER'S AD WAS VIEWED BY AGE AND GENDER
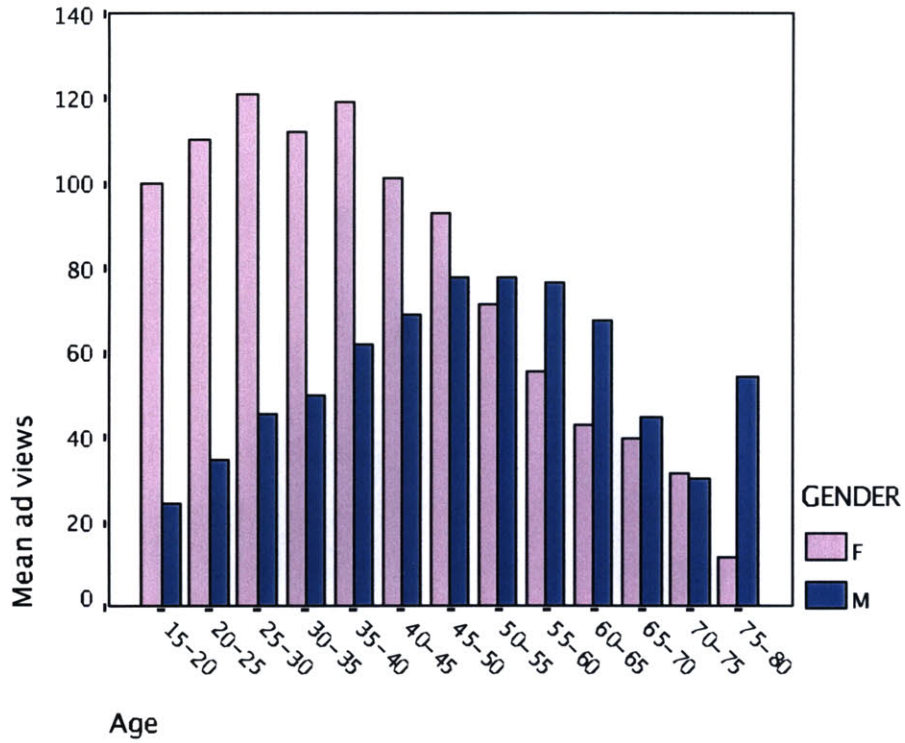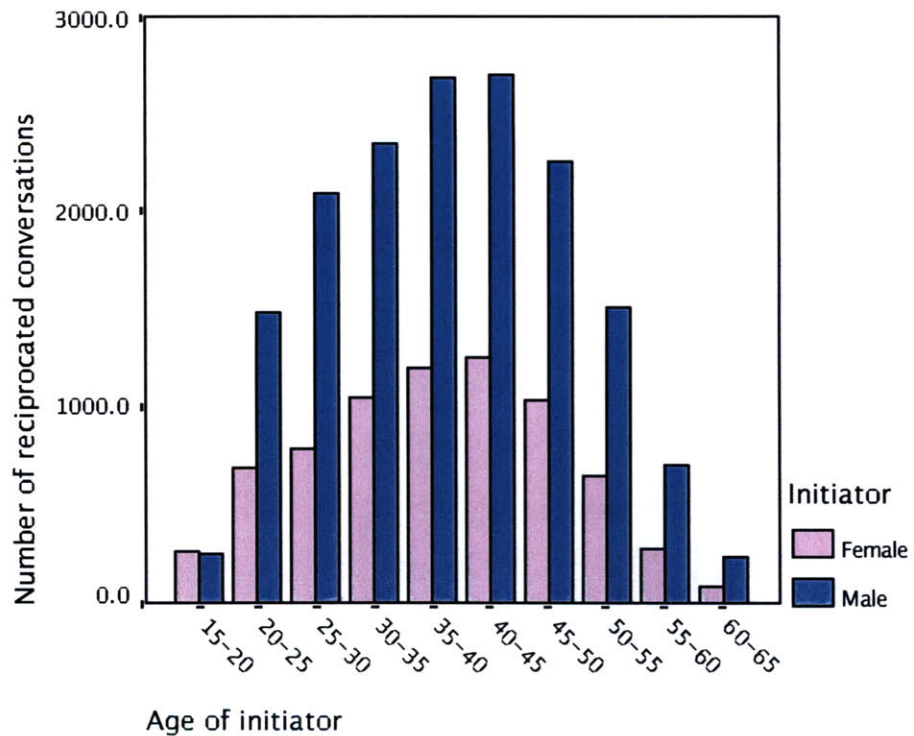

B.2 MEAN NUMBER OF MESSAGES RECEIVED BY AGE AND GENDER
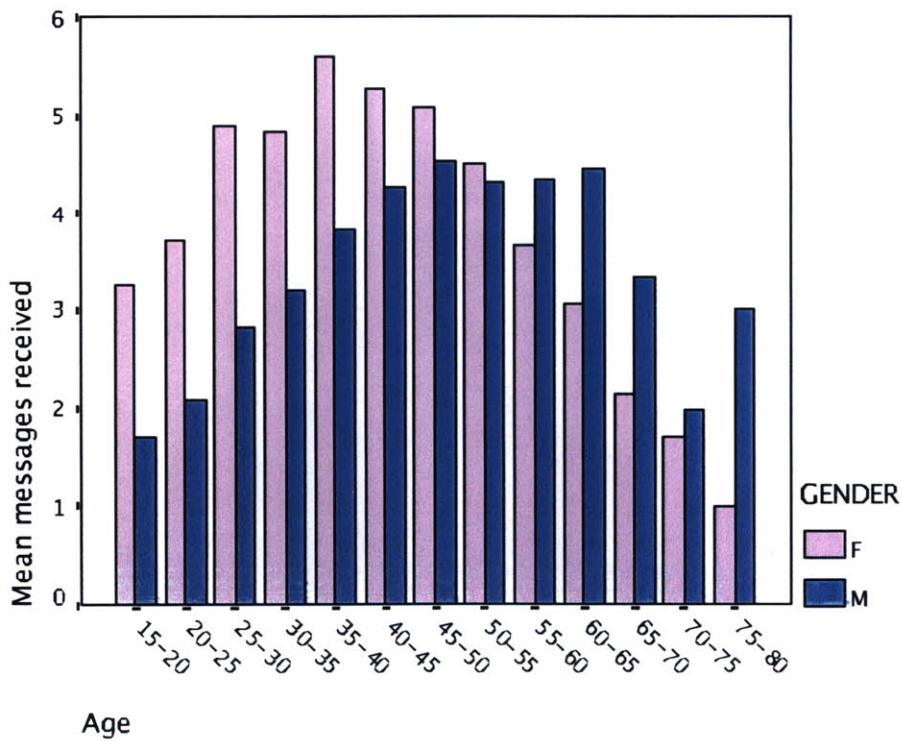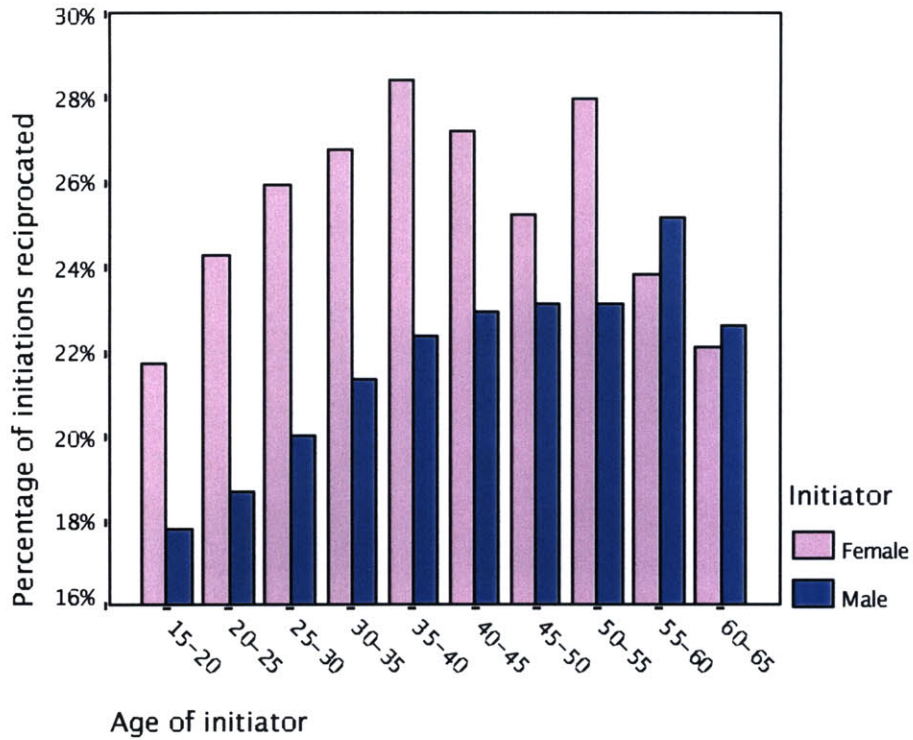
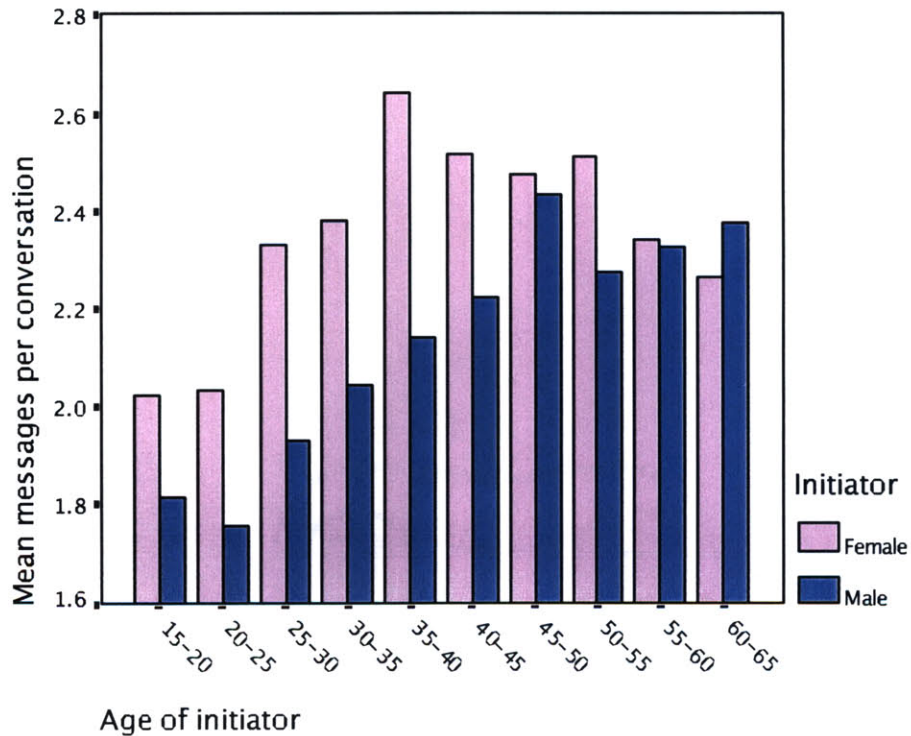## B.3 NUMBER OF CONVERSATIONS INITIATED BY AGE AND GENDER
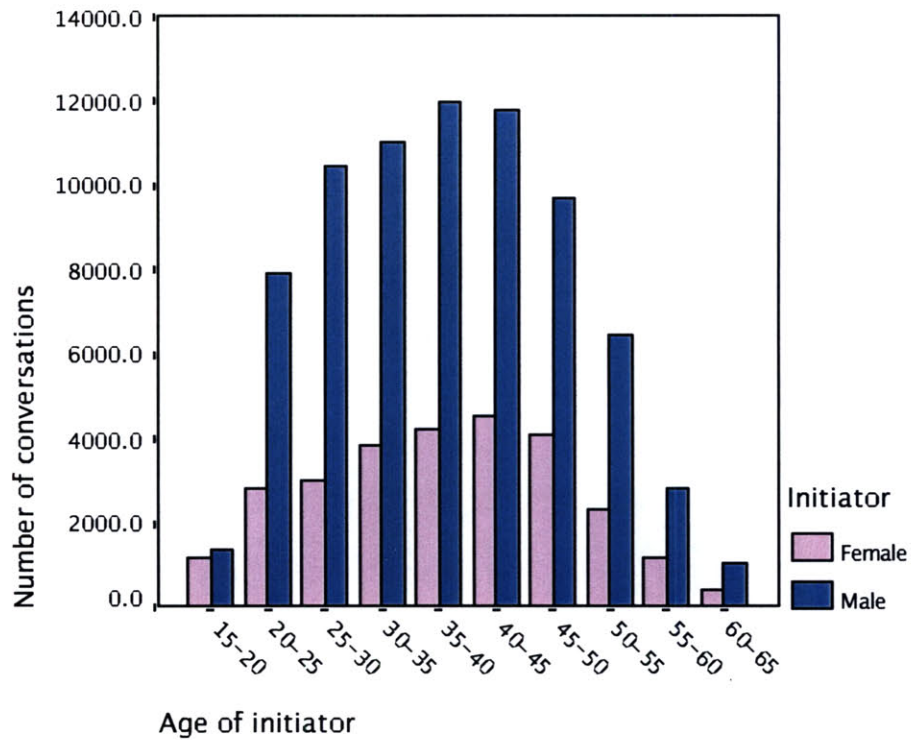
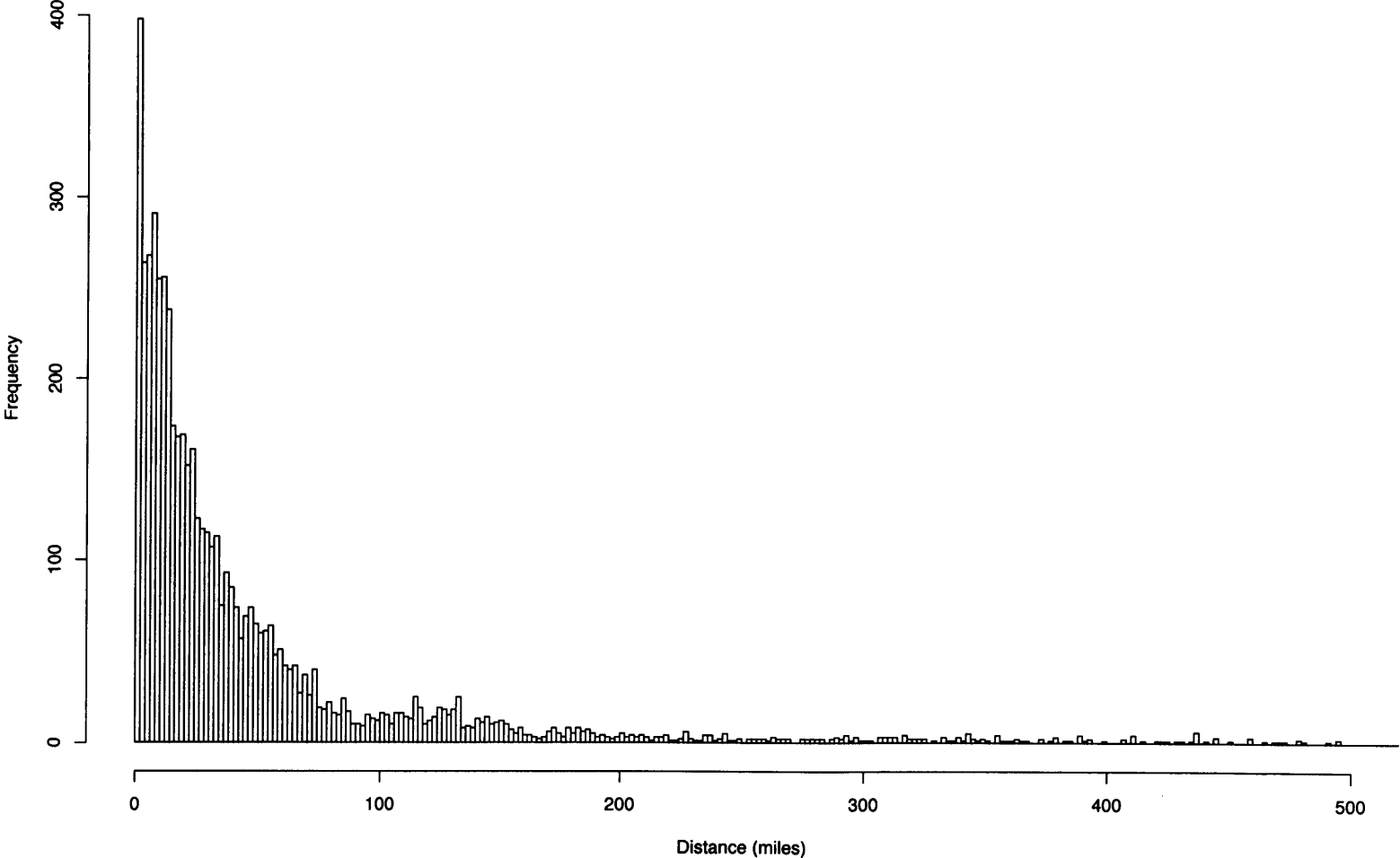## B.4 NUMBER OF RECIPROCATED CONVERSATIONS INITIATED BY AGE AND GENDER

B.5 PERCENTAGE OF INITIATIONS RECIPROCATED BY AGE AND GENDER OF INITIATOR

B.6 MEAN MESSAGES PER CONVERSATION BY AGE AND GENDER OF INITIATOR

B.7 HISTOGRAM OF DISTANCE BETWEEN MEMBERS WHO EXCHANGED ONE OR MORE MESSAGES

B.8 RESPONSE RATES BY GENDER OF INITIATOR

Percent	Initiator (A) gender	Recipient (B) gender	Percent reciprocated	Mean msgs by A	Mean msgs by B
26.7	F	M	25.1	1.82	0.57
73.3	M	F	20.6	1.66	0.45

B.9 REPONSE RATES BY GENDER AND AGE-IN-PREFERRED-RANGE

Percent of contacts	Initiator (A) gender	Recipient (B) gender	Init's age matches Recip's prefs	Recip's age matches Init's prefs	Percent no reply from B	Avg msgs from Init[1]	Avg msgs from Recip[1]
0.71	F	M	No	No	76.83	1.70 (2.52)	0.44 (1.92)
3.84	F	M	No	Yes	78.36	1.67 (2.76)	0.43 (2.02)
3.15	F	M	Yes	No	71.26	1.89 (2.57)	0.61 (2.14)
18.76	F	M	Yes	Yes	73.14	1.85 (2.80)	0.60 (2.26)
2.80	M	F	No	No	83.77	1.53 (2.80)	0.31 (1.96)
20.40	M	F	No	Yes	84.84	1.52 (2.87)	0.29 (1.96)
3.35	M	F	Yes	No	70.30	1.81 (2.62)	0.64 (2.15)
45.04	M	F	Yes	Yes	75.45	1.73 (2.89)	0.51 (2.11)

1. Initial numbers show mean over all cases. Numbers in parentheses show mean in the cases where B replies.

B.10 LINEAR REGRESSION MODELS TO PREDICT LOG$_e$(*MESSAGES RECEIVED PER DAY*)

For male and female users separately. *B* indicates the unstandardized coefficient, whereas *Beta* is standardized.

	Men			Women		
	B (std. err.)	*Beta*	*Signif.*	*B* (std. err.)	*Beta*	*Signif.*
Avg. msgs sent per day	5.627 (0.126)	0.393	***	6.203 (0.129)	0.404	***
Never married	0.0326 (0.016)	0.025	*	0.0344 (0.015)	0.026	*
Wants children	– 0.0288 (0.015)	– 0.019	n.s.	– 0.00249 (0.015)	– 0.002	n.s.
Has children	0.0382 (0.014)	0.029	**	0.0325 (0.013)	0.026	*
Is Christian	0.0478 (0.012)	0.036	***	0.0212 (0.011)	0.016	n.s.
Is Caucasian	0.110 (0.021)	0.046	***	0.130 (0.019)	0.060	***
Has photo	0.0296 (0.011)	0.023	**	0.0966 (0.011)	0.074	***
Age	0.0174 (0.001)	0.147	***	0.00735 (0.001)	0.063	***
Educational level	0.0751 (0.010)	0.066	***	0.0662 (0.010)	0.057	***
Phys. attractiveness	0.112 (0.022)	0.049	***	0.173 (0.020)	0.078	***
Small build	0.0106 (0.022)	0.004	n.s.	0.0292 (0.014)	0.020	*
Athletic build	0.00559 (0.013)	0.004	n.s.	0.0283 (0.019)	0.013	n.s.
Heavy build	– 0.0861 (0.020)	– 0.040	***	– 0.161 (0.015)	– 0.096	***
Height	– 0.00344 (0.004)	– 0.008	n.s.	– 0.00310 (0.004)	– 0.007	n.s.
Smokes	– 0.0532 (0.011)	– 0.044	***	– 0.0389 (0.010)	– 0.034	***
Drinks	0.0316 (0.016)	0.018	*	0.0598 (0.016)	0.032	***
Constant	– 4.518 (0.291)			– 4.053 (0.266)		
R-squared	0.213			0.209		

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

B.11 RESIDUAL DIAGNOSTICS FOR MEN'S MODEL IN B.10

Histogram of Standardized Residual

Dependent Variable: ln(messages received per day)

Std. Dev = 1.00
Mean = 0.00
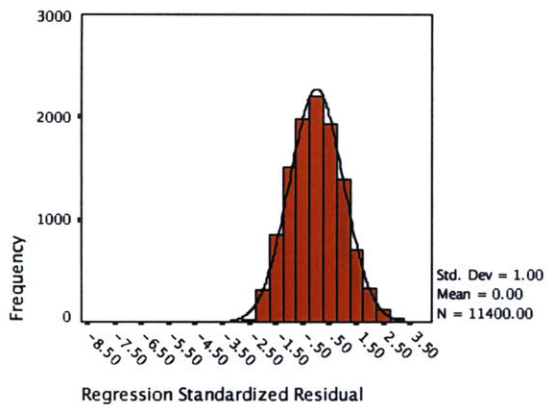N = 10305.00
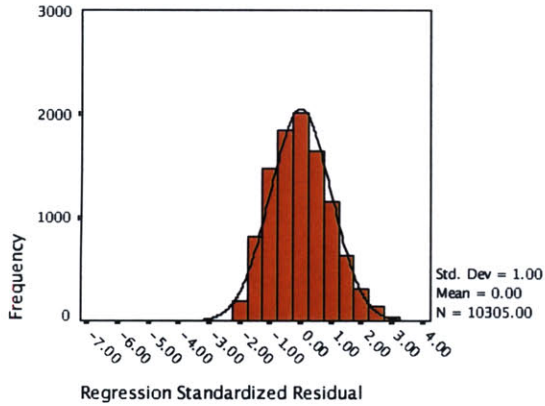
Normal P-P Plot of Standardized Residual

Dependent Variable: ln(messages received per day)
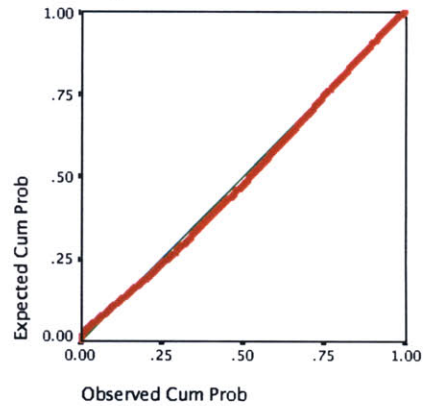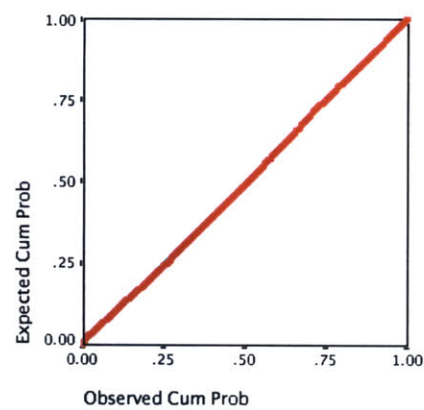

B.12 RESIDUAL DIAGNOSTICS FOR WOMEN'S MODEL IN B.10

Histogram of Standardized Residual

Dependent Variable: ln(messages received per day)

Std. Dev = 1.00
Mean = 0.00
N = 11400.00

Normal P-P Plot of Standardized Residual

Dependent Variable: ln(messages received per day)

B.13 PLOT WITH CONNECTIONS OF ACTIVE USERS BY NUMBER OF CHILDREN AND MARITAL STATUS

2845-M9