It's no secret
Measuring the security and reliability of authentication via 'secret' questions

Stuart Schechter
Microsoft Research
stus@microsoft.com

A. J. Bernheim Brush
Microsoft Research
ajbrush@microsoft.com

Serge Egelman
Carnegie Mellon University
egelman@cs.cmu.edu

Abstract

All four of the most popular webmail providers – AOL, Google, Microsoft, and Yahoo! – rely on personal questions as the secondary authentication secrets used to reset account passwords. The security of these questions has received limited formal scrutiny, almost all of which predates webmail. We ran a user study to measure the reliability and security of the questions used by all four webmail providers. We asked participants to answer these questions and then asked their acquaintances to guess their answers. Acquaintances with whom participants reported being unwilling to share their webmail passwords were able to guess 17% of their answers. Participants forgot 20% of their own answers within six months. What's more, 13% of answers could be guessed within five attempts by guessing the most popular answers of other participants, though this weakness is partially attributable to the geographic homogeneity of our participant pool.

1. Introduction

The four largest webmail providers – AOL, Google, Microsoft, and Yahoo! – all use personal (a.k.a. 'secret') questions to authenticate account holders who are unable to login using their passwords. While other web services may authenticate users who have forgotten their passwords via their email addresses, webmail services cannot always do so; many of their users employ their accounts as a primary email address and may not have another dependable email account for use as a backup authenticator. These same users often rely on their webmail addresses as backup authenticators for other services, raising the consequences should their webmail accounts' authentication mechanisms fail.

Despite the consequences of authentication failures, the four largest webmail providers require only one question be answered in order to reset an account's password. Concerns over the security of these questions abound, in part, because webmail is so popular; the top two webmail services each claim a quarter of a billion active users [7], [9]. Public awareness of the potential weaknesses of personal authentication questions reached new heights when 2008 Republican vice presidential nominee Sarah Palin's Yahoo! Mail account was compromised via her question [2].

In fact, prior research suggests that a single personal question is not a sufficiently secure authenticator. In two different studies, one in 1990 [17] and another in 1996 [10], participants were asked personal authentication questions and those they were close to – spouses, family members, or close friends – were able to guess 33%-39% of their answers. These studies also addressed the memorability of questions; participants forgot 20%-22% of their own answers within three months.

Given these statistics, why do webmail providers still authenticate users by asking a single personal question? Perhaps they have used the past twelve years to develop a generation of questions with answers that are easier to remember and harder for others to guess. Or maybe earlier research should be disregarded because, while answers were vulnerable to guessing by trusted significant others, less trusted acquaintances would be far less likely to guess the correct answers.

To quantify the security and reliability of personal authentication questions as they are used today, we examined the real-world questions in use as of March 2008 by the top four webmail providers. We invited participants to our laboratory in pairs, asked them these personal questions, and then asked them to guess their partners' answers. We extend prior research by measuring the security of these questions against guessing not just by significant others, but by *untrusted* acquaintances as well. We also examine the vulnerability of these questions to *statistical guessing* attacks, which

identify the most popular answers to each question and try each one until no more guesses are allowed.

For those participants who brought partners who they would not trust with their Hotmail password, we found that these partners could still guess an alarming 17% of their answers. Many answers could be guessed without even knowing the participant. From the geographically-homogenous set of participants in our laboratory study, 13% of their answers could be guessed within five attempts using statistical guessing. For user-written question/answer pairs, we categorized roughly 25% as vulnerable to family members, friends, or coworkers and another 15% as guessable within five tries with no knowledge of the victim.

2. Background and Related Work

Personal authentication questions have received a great deal of attention from the popular press. The most recent burst in coverage came as the result of the revelation that 2008 Republican vice presidential nominee Sarah Palin's Yahoo! account had been compromised by someone who researched the answer to the question *Where did you meet your spouse?* [2], [6]. Though the Palin story focused on the security of these questions, others have focused on their reliability as an authenticator and the plight of those who cannot get into their accounts [14].

The Palin incident came only months after coverage of a paper by Ariel Rabkin who examined the questions used by twenty bank websites [11]. He manually categorized questions he believed to be ambiguous, not applicable to over 15% of the general public, not memorable, easily guessable with no knowledge of the victim, or easily guessable with minimal knowledge of the victim. However, he did not actually quantify the level of vulnerability that resulted from any given question.

The use of personal questions for authentication was studied by Zviran and Haga in 1990 [17]. They examined how well others might be able to guess the answer to users' personal authentication questions, but focused on guessing by "significant others," the great majority of whom were participants' spouses (77%). The remainder were close friends (17%), siblings (4%), and parents (2%). Zviran and Haga did not report whether they collected their data electronically or on paper, or whether they compared answers manually or algorithmically. They only allowed one guess in both the recall and guessing phases. Partners in their study guessed 33% of participants' answers, which was quite similar to the 34% of answers guessed by spouses in our study (see the first data column of Table 6). Participants in their study recalled 78% of their answers after three months, which was also similar to the figures in our study (80%, which is expressed as a recall failure rate of 20% in the second data column of Table 4).

Podd *et al.* conducted a similar study in 1996, and found similar recall rates (80%) and higher guessing rates (39.5%) [10]. They, too, focused on partners who were significant others.

Many of the 20 questions Zviran and Haga used were likely to have a small set of common answers—e.g. *favorite color*, *favorite class in high school*, and *favorite flower*. Such small answer spaces may have been acceptable to Zviran and Haga as they proposed the use of multiple questions to reduce false authentications and rejections. To our knowledge, no previous research has explored the vulnerability of personal questions to statistical attacks: those that walk down a list of the most popular answers for a target population. Statistical guessing attacks could be refined further by examining answer popularity as a function of the language of the account holder, geographic locale, or other traits discernable to an attacker.

Despite earlier findings and unanswered questions about their security, personal questions have been adopted for use as a backup authentication mechanism by all of the top four webmail providers (as identified by Hitwise [8]): AOL, Google (Gmail), Microsoft (Hotmail), and Yahoo!. All rely on a single question, though some may also verify the user's zip code. Google will not allow a password to be reset until an account has been inactive for a period of time. While many other sites also use personal questions for backup authentication, webmail services are uniquely dependent on them because they cannot assume their users have an alternate email address as a backup authenticator.

Google also lets users opt to write their own security question, which some have speculated is more secure than relying on standardized questions [15]. While we refer to these as user-written questions, others have called them 'open' questions [5]. Toomim *et al.* have investigated using user-written questions for authenticating members of social groups, as motivated by a scenario in which an individual wants to share photos with friends who were at the same party [16]. They investigated the security of these questions by offering rewards for correct answers on Mechanical Turk. However, the simulated attackers in their study were at a disadvantage compared to real attackers: they had no contextual knowledge about who had written the question or what it was intended to protect.

3. Study recruitment and methodology

To study the reliability and security of personal questions, we ran a laboratory study over four separate days between March 22 and June 23, 2008, with a follow-up study in September and October. The cohorts assigned to each day are shown in in Table 2a. The study encompassed both the personal questions used by Windows Live's password-reset workflow and the questions used by the top four webmail services.

3.1. Participant recruitment

Our recruiting team selected participants from a larger pool of potential participants they maintain for all studies at Microsoft. The pool contains members of the general public who had been recruited via public events, lotteries, and our website. We required that participants speak English as their primary language and not be employed by Microsoft.

Our recruiters selected a balance of men and women; 64 participants were male and 66 female. The recruiters also selected participants with a diversity of ages and professions. While the professions are too numerous to list, the age ranges are broken down in Table 2b.

Participants in the first three cohorts were required to be Hotmail users for at least three months and to access their account at least three times a week. The great majority of participants (83%) had been using their Hotmail account for at least four years, as detailed in Table 2d.

After reaching one qualified participant, our recruiters would ask if the participant had a coworker, friend, or family member who might also be qualified for the study. Recruiters then interviewed potential partners to ensure they met our requirements. All participants were required to have partners and the categories of relationships between participants and their partners are broken down in Table 2c.

3.2. Initial laboratory visit

We scheduled participants for a two-hour visit to perform the tasks summarized in Table 1.

Participants in each session were split into groups and placed into different rooms such that no two partners were in the same room. Each partner was placed at a computer. We seated participants sufficiently far from each other to ensure that their screens, on which their answers might appear while being typed, could not be seen by others. All questions were asked using web survey software, though participants were required to be on-site to prevent collusion.

Table 1. Order of laboratory visit tasks

1) Move to room separate from partner
2) Answer demographic questions
3) Authenticate to Hotmail using personal question (cohorts 1-3)
4) Answer personal questions for top four webmail services
5) Describe relationship with partner
6) Guess partner's answers to personal questions
7) Attempt to recall answers to own personal questions
8) Second chance to guess partner's questions using online research (cohorts 2-4)

3.2.1. Authentication to Hotmail. We explained to participants how personal questions could be used to reset the passwords participants' used to login to Hotmail. We asked the 116 participants in the first three cohorts (those selected to be Hotmail users) to attempt to answer their personal question. We asked them only to authenticate (provide the answer to their question) and not to actually reset their password if successful.

3.2.2. Initial answers to personal questions. We then asked all 130 participants to answer all of the personal questions in use by the top four webmail services. We told participants that we would ask the same questions later to determine how well they remembered the answers. We offered two prizes (an XBOX 360 and a Zune digital music player) and gave participants a virtual lottery ticket for each question they both answered and later recalled.

We randomized the question order for each participant. We asked participants to mark questions they were either unable or unwilling to answer. We instructed participants that capitalization, punctuation, and spaces would be ignored when comparing answers.

We anticipated participants might try to increase their chance of recalling their answers by providing the same answer for all questions. We added a rule that eliminated rewards for recalling the same answer numerous times. We also feared that if participants anticipated being asked to recall their questions again at a future date, they might record their answers following the study session. We thus asked participants to recall their answers at the end of their session and ran the lottery for the laboratory session prizes based on these recollections. We did not inform participants that we would follow-up to test their recollections in the future.

After participants had been asked all of the questions used by the top four webmail services, we asked them what they would choose if they could write their own question. We also asked them to answer the question they wrote.

date of first visit	# ppts in study main	recall
March 22	40	15
April 26	44	20
May 31	32	14
June 23	14	0
Total	130	49

(a) Cohorts

age group	participants	
< 18	2	(2%)
18–25	28	(22%)
26–35	51	(39%)
36–55	31	(24%)
55+	18	(14%)

(b) Age groups

relationship to partner	participants	
Spouse	18	(14%)
Relative	23	(18%)
Fiance/SO	4	(3%)
Friend	51	(39%)
Coworker	32	(25%)
Other	2	(2%)

(c) Relationships

webmail account age	participants	
< 6 months	6	(5%)
½–1 year	4	(3%)
1–4 years	10	(9%)
> 4 years	96	(83%)

(d) Webmail account ages

Table 2. Demographics

3.2.3. Guessing by acquaintances. We asked participants to describe their relationship with their partner and asked them whether they would trust their partner with their Hotmail password. Then we asked them to guess their partners' answers. As before, we presented the questions in random order and rewarded success with an increased opportunity to win one of our prizes, though we could not tell participants which answers were correct. We allowed participants to guess up to five times by placing guesses on separate lines. We restricted participants from communicating answers to each other by asking them to turn off their mobile devices ("as a courtesy to others"), isolating them in separate rooms, and monitoring their behavior.

After running the first cohort of the study (40 participants), we discovered that many participants weren't guessing as hard as we had hoped. Most were providing at most one guess per answer and none appeared to be performing any online research. We thus gave the 90 participants in the three remaining cohorts (cohorts 2–4) a second opportunity to guess their partners' answers. In this second guessing round, we encouraged them to use search engines and social networking sites to research the answers to their partners' questions. We also told them that this was the last task of the study in hopes that they might feel less rushed.

3.3. Reliability (memorability) follow-up

To determine how well participants remembered the answers to the personal questions we had asked, we followed up with them between September 5 and October 31. Of the 116 participants in the first three cohorts, we contacted all 87 who had consented to receive emails from us and 49 volunteered to participate.

We used a custom-built web tool to ask participants to recall the answers to the questions that they had chosen to answer in the laboratory study. For each question, we allowed them to respond as many times as they liked until they either correctly recalled their original answer or chose to move onto the next question. Answers were judged as correct recollections if they differed from the original only in the use of white space, punctuation, and capitalization. This was the strictest of the comparison algorithms we wanted to examine. By only acknowledging a participant's answer as correct if it met the strictest requirements, we could later test how less strict algorithms would have increased recall rates and reduced the number of attempts required.

To encourage participants to do their best at recalling their original answers we offered all participants a new incentive, again based on the percentage of answers they recalled. The top quartile received an Amazon.com gift card worth $15, the second quartile received one worth $10, the third $5, and the last quartile received no performance-based gratuity. In addition, all participants received some form of base gratuity just for participating; some participants were offered a software gratuity for completing the recall task along with a separate study, whereas others were offered a $10 Amazon.com gift certificate for completing the recall study alone.

3.4. Limitations

While we provided incentives for participants to answer questions as if they were setting up their account, and to guess their partners' answers as best they could, they may not have done so.

Some individuals may be more invested in picking a memorable and secure question/answer pair when setting up a real account than when in the lab [13]. Others may discount the need for secure and reliable backup authentication when setting up a real account, but feel obligated to help researchers when in the lab.

While participants in the laboratory had to guess the answers to all questions during a limited amount of time, a real attacker need only answer one question to compromise an account and may invest as much time as he or she wishes. Thus, our estimates of the abilities of acquaintances to guess answers are likely to underestimate their true potential.

Table 3. Answer comparison algorithms

algorithm	forgot within 3–6 months	guessed by partner	guessed by partner, broken down by *would you trust your partner with your Hotmail password?*		
			no	some circumstances	yes
equality	256/1070 (23.9%)				
substring	240/1070 (22.4%)	588/2870 (20.5%)	110/662 (16.6%)	146/942 (15.5%)	332/1266 (26.2%)
distance	213/1070 (19.9%)	628/2870 (21.9%)	115/662 (17.4%)	162/942 (17.2%)	351/1266 (27.7%)

The *equality* algorithm could not be run on partners' guesses because our survey tool represented all guesses as a single concatenated string (see Section 4).

4. Answer comparison algorithms

In total, 130 participants initially provided 2,874 answers and 49 participated in the follow-up study and tried to recall 1,074 of those answers. We needed an algorithm for determining whether a recollection, or partner's guess, sufficiently matched the original. We tested three different algorithms.

For all algorithms, we removed all non-alphanumeric characters and forced letters into lower case. When counting the number of attempts to recall an answer, we did not count repetitions of the same guess.[1] Attackers learn nothing by being able to repeat a guess, whereas account holders, who may repeat the same answer thinking they previously mistyped it, will not be penalized for this mistake.

The first algorithm, simple *equality*, compares the resulting simplified strings character for character. This is the algorithm that was used, during the memorability follow-up study, to provide participants with feedback as to whether they had recalled their answers correctly.

Unfortunately, we could not use the equality algorithm for examining partners' guesses due to an artifact of our study. The Illume survey software we used to collect the guesses participants provided for their partners' answers fails to store carriage returns, which we had asked participants to use to separate their guesses.

To address this problem our second algorithm, the *substring* algorithm, treated a guess as valid if it contained a substring that matched the original answer, as suggested by Toomim *et al.* [16].

The final algorithm we tested was the Levenshtein edit *distance* algorithm with two modifications. First, we reduced the cost of transpositions of two characters ('swapped'→'sawpped') from two to one. This reduces the cost of this very common typo to be equal to that of a single mistyped character. Second, we removed the cost of extra characters at the beginning or end of the guess, to adjust for the artifact that all guess strings were concatenated together. We allowed one error (an edit distance cost of one) for every five characters in the original answer.

Table 3 illustrates the performance of each algorithm over all of the questions.

Moving from the substring algorithm to the distance algorithm reduces the number of answers *forgotten* (not recalled within 5 attempts) by 2.5% as a percent of total answers, from 22.4% to 19.9%. This represents a 11.3% reduction from the answers deemed forgotten by the substring algorithm.

Alas, moving from the substring algorithm to the distance algorithm also increased the percentage of answers guessed by participants' partners by 1.4%. That's a 6.8% relative increase over the percent guessed using the substring algorithm. However, when we closely analyzed the answers that had been reclassified from not guessed to correctly guessed, we were convinced the trade-off was well worth it. In 34 of the forty cases where a guess was treated as incorrect by the substring algorithm but correct by the distance algorithm (80%), the guessing partner clearly knew the correct answer: the difference was a one character typing error that an attacker could easily fix with a second guess. In four of the remaining six cases, it was clear that the partner knew the answer but excluded a few characters, such as entering a city but excluding a two character state suffix. In only two cases did manual inspection fail to reveal convincingly that the partner knew the answer. Those two cases represent less than a 0.1% increase in total answers guessed over the substring algorithm.

Given that the benefit of the distance algorithm appeared to greatly outweigh its cost, we used it for the duration of the study and recommend a variant for real-world deployment. In such a deployment, the length of the guess should be truncated (as a function of the original answer length) so that an attacker cannot concatenate multiple guesses together.

5. Results

We briefly cover the results for participants who tried to authenticate to their Hotmail accounts using their personal question, then examine the results from our data on the top four webmail services' questions.

1. We first learned of this heuristic from Charlie Kaufman.

5.1. Real-world memorability results

While we asked all 116 participants in the first three cohorts to try to reset their password using their personal question, not all accounts had a question configured. Furthermore, an answer alone was not sufficient to authenticate: a zip code previously associated with the account was also required.

A total of 99 participants reported being asked to provide the answer to their personal question. Only 43 (43%) reported being able to successfully provide the correct answer and their zip code. The majority, 56 (57%) could not reset their password and reported being unable to remember either the answer or the zip code they had provided when they set up the account.

When asked why they had trouble authenticating, 75% participants suspected they may have been unable to answer their personal question and 31% reported that they may have been unable to recall the zip code they had previous provided. A surprising 13% of participants suspected that the reason they could not answer their personal question was because they had intentionally provided a bogus answer when setting up their account.

5.2. Willingness to answer

The results for all questions used by the top four webmail services[2] (as of March, 2008) are summarized in Table 4. The questions appear in the order in which the webmail services present them to the user.

The first data column of Table 4 shows the number and percentage of participants who opted to answer each question. We excluded all answers in which participants expressed being uncomfortable, unwilling, or unable to provide an original answer. While we had prescribed a method of indicating a non-answer (*n/a* for not applicable and *n/c* for not comfortable), we manually identified numerous other indicators used by participants, such as "not willing", "unknown", and "don't have one", and treated them as non-answers as well.

For three of the four services (AOL, Microsoft, and Yahoo!), participants opted to answer their questions between 81% and 85% of the time. All participants opted to answer at least one of Yahoo!'s questions and only one participant opted not to answer any of AOL's or Microsoft's questions.

In contrast, participants opted to answer each of Google's questions an average of 50% of the time and 14 (11%) opted not to answer any of them. Google lets users choose to write their own personal question, the implication of which are examined in Section 5.6.

2. One question used by Microsoft, *Name of first pet*, is excluded due to a data collection error documented in Appendix A.

5.3. Reliability (memorability)

The second data column of Table 4 shows the number and percentage of participants who answered each question, but who were unable to recall their answer within five guesses during the follow-up study. For those who did recall their answer within five guesses, 76% did so on the first guess. A detailed breakdown of the number of guesses required is in Table 8 in the Appendix.

One participant answered all questions with "password", which he was able to remember when asked to recall his answers at the end of the laboratory session. However, during the follow-up study he had forgotten that he had done this and so he failed to answer all questions. This individual was responsible for one of the answers forgotten in every row of the 'forgot' column. We opted not to remove this contribution because this may be a real-world mechanism for coping with these questions, even if it proved ineffective in this case.

Among the questions with answers forgotten 25% of the time or more, which appear in boldface in the second column of Table 4, all but one fall into two categories: preferences and ID numbers. Preferences may be hard to remember because a participant's choice of childhood hero, historical person, song, film, or pastime may be subject to whims of the moment. ID numbers, such as frequent flyer and library card numbers, may not have been stored in memory to start with. Remembering the correct frequent flyer number may be particularly difficult if one has many frequent flyer accounts or if one's favorite airline goes bankrupt.

5.4. Security against statistical guessing

The third data column of Table 4 shows the vulnerability of answers to a statistical guessing attack. An answer is deemed vulnerable to this attack if it is among the five most popular answers provided by *other* participants (excluding the participant's partner). In other words, we compute the five most popular answers for all participants except the participant who answered the question and that participant's partner, break ties randomly, and then mark an answer as statistically guessable if it matches one of those five answers. We have highlighted in boldface those questions for which more than 10% of answers were statistically guessable.

Table 4. Questions used by the top four webmail service providers

	answered *130 ppts*	forgot within 3–6 months *49 ppts*	statistically guessable *130 ppts*	guessed by partner *130 ppts*	guessed by partner broken down by *would you trust your partner with your Hotmail password?* no *32 ppts*	under some circumstances *42 ppts*	yes *56 ppts*	2nd round guess improvement *90 ppts*
AOL								
What is your pet's name?	93 (72%)	3/40 (8%)	0/93 (0%)	45/93 (48%)	8/20 **(40%)**	11/28 (39%)	26/45 (58%)	1/61 (2%)
Where were you born?	129 (99%)	1/48 (2%)	19/129 **(15%)**	60/129 (47%)	14/31 **(45%)**	14/42 (33%)	32/56 (57%)	6/50 **(12%)**
What is your favorite restaurant?	117 (90%)	11/44 **(25%)**	7/117 (6%)	20/117 (17%)	2/28 (7%)	6/37 (16%)	12/52 (23%)	1/77 (1%)
What is the name of your school?	96 (74%)	6/35 (17%)	22/96 **(23%)**	27/96 (28%)	6/20 **(30%)**	7/33 (21%)	14/43 (33%)	4/71 (6%)
Who is your favorite singer?	101 (78%)	5/34 (15%)	1/101 (1%)	12/101 (12%)	2/24 (8%)	2/36 (6%)	8/41 (20%)	0/83 (0%)
What is your favorite town?	115 (88%)	8/43 (19%)	34/115 **(30%)**	33/115 (29%)	5/26 (19%)	11/38 (29%)	17/51 (33%)	2/66 (3%)
What is your favorite song?	94 (72%)	14/33 **(42%)**	3/94 (3%)	4/94 (4%)	0/24 (0%)	1/30 (3%)	3/40 (8%)	2/88 (2%)
What is your favorite film?	114 (88%)	15/42 **(36%)**	11/114 (10%)	18/114 (16%)	2/24 (8%)	7/38 (18%)	9/52 (17%)	4/82 (5%)
What is your favorite book?	100 (77%)	8/35 (23%)	19/100 **(19%)**	12/100 (12%)	2/22 (9%)	2/36 (6%)	8/42 (19%)	3/84 (4%)
Where was your first job?	125 (96%)	10/48 (21%)	10/125 (8%)	26/125 (21%)	4/30 (13%)	7/39 (18%)	15/56 (27%)	1/74 (1%)
Where did you grow up?	127 (98%)	3/48 (6%)	22/127 **(17%)**	61/127 (48%)	15/31 **(48%)**	15/40 (38%)	31/56 (55%)	4/47 (9%)
Total	1211 (85%)	84/450 (19%)	148/1211 (12%)	318/1211 (26%)	60/280 **(21%)**	83/397 (21%)	175/534 (33%)	28/783 (4%)
Google								
What is your primary frequent flyer number?	29 **(22%)**	5/10 **(50%)**	0/29 (0%)	0/29 (0%)	0/8 (0%)	0/9 (0%)	0/12 (0%)	0/90 (0%)
What is your library card number?	39 **(30%)**	8/16 **(50%)**	0/39 (0%)	0/39 (0%)	0/8 (0%)	0/19 (0%)	0/12 (0%)	0/90 (0%)
What was your first phone number?	93 (72%)	6/36 (17%)	0/93 (0%)	9/93 (10%)	1/20 (5%)	3/35 (9%)	5/38 (13%)	0/82 (0%)
What was your first teacher's name?	93 (72%)	7/31 (23%)	6/93 (6%)	1/93 (1%)	0/18 (0%)	1/32 (3%)	0/43 (0%)	0/89 (0%)
Total	254 (49%)	26/93 **(28%)**	6/254 (2%)	10/254 (4%)	1/54 (2%)	4/95 (4%)	5/105 (5%)	0/351 (0%)
Microsoft								
Mother's birthplace	121 (93%)	6/44 (14%)	12/121 (10%)	34/121 (28%)	5/29 (17%)	8/40 (20%)	21/52 (40%)	5/69 (7%)
Best childhood friend	120 (92%)	8/44 (18%)	1/120 (1%)	23/120 (19%)	3/28 (11%)	4/37 (11%)	16/55 (29%)	1/79 (1%)
Favorite teacher	105 (81%)	8/38 (21%)	0/105 (0%)	5/105 (5%)	1/23 (4%)	3/33 (9%)	1/49 (2%)	1/88 (1%)
Favorite historical person	106 (82%)	15/40 **(38%)**	27/106 **(25%)**	13/106 (12%)	4/24 (17%)	3/35 (9%)	6/47 (13%)	0/83 (0%)
Grandfather's occupation	99 (76%)	12/37 **(32%)**	13/99 **(13%)**	12/99 (12%)	4/22 (18%)	4/30 (13%)	4/47 (9%)	1/82 (1%)
Total	551 (85%)	49/203 (24%)	53/551 (10%)	87/551 (16%)	17/126 (13%)	22/175 (13%)	48/250 (19%)	8/401 (2%)
Yahoo!								
Where did you meet your spouse?	80 (62%)	8/36 (22%)	10/80 **(13%)**	23/80 (29%)	4/21 (19%)	4/20 (20%)	15/39 (38%)	3/76 (4%)
What was the name of your first school?	116 (89%)	4/43 (9%)	1/116 (1%)	23/116 (20%)	2/24 (8%)	6/38 (16%)	15/54 (28%)	3/74 (4%)
Who was your childhood hero?	97 (75%)	16/33 **(48%)**	27/97 **(28%)**	10/97 (10%)	6/24 **(25%)**	2/30 (7%)	2/43 (5%)	4/87 (5%)
What is your favorite pastime?	118 (91%)	15/44 **(34%)**	23/118 **(19%)**	28/118 (24%)	5/29 (17%)	6/39 (15%)	17/50 (34%)	3/70 (4%)
What is your favorite sports team?	100 (77%)	3/38 (8%)	57/100 **(57%)**	47/100 (47%)	7/22 **(32%)**	15/35 (43%)	25/43 (58%)	5/66 (8%)
What is your father's middle name?	108 (83%)	2/42 (5%)	5/108 (5%)	17/108 (16%)	3/26 (12%)	2/35 (6%)	12/47 (26%)	1/81 (1%)
What was your high school mascot?	109 (84%)	3/42 (7%)	18/109 **(17%)**	34/109 (31%)	4/26 (15%)	10/36 (28%)	20/47 (43%)	7/67 **(10%)**
What make was your first car or bike?	126 (97%)	3/46 (7%)	31/126 **(25%)**	31/126 (25%)	6/30 **(20%)**	8/42 (19%)	17/54 (31%)	0/68 (0%)
What is your pet's name?	93 (72%)	3/40 (8%)	0/93 (0%)	45/93 (48%)	8/20 **(40%)**	11/28 (39%)	26/45 (58%)	1/61 (2%)
Total	947 (81%)	57/364 (16%)	172/947 (18%)	258/947 (27%)	45/222 **(20%)**	64/303 (21%)	149/422 (35%)	27/650 (4%)
Total for all webmail sites	2870 (79%)	213/1070 (20%)	379/2870 (13%)	628/2870 (22%)	115/662 (17%)	162/942 (17%)	351/1266 (28%)	62/2124 (3%)

The column labeled *answered* contains the number of participants who opted to answer each question out of the 130 who participated. The column labeled *forgot* contains the number of participants who, during the follow-up study, were unable to recall the correct answers (as judged by the distance algorithm) within five tries. Percentages are based on the number of questions answered and those above 25% are in bold. A participant's answer was deemed *statistically guessable* if it was among the five most common answers chosen by all other participants (excluding the participant's partner). Those above 10% are highlighted in boldface. The column labeled *guessed by partner* contains the number of answers guessed by the partner within five tries (using the distance algorithm). In the breakdown of partners not trusted with the answering participant's Hotmail password, we highlight those questions guessed by more than 20% these untrusted partners.

The guesses in the *guessed by partner* columns included those given in a second round of research-based guessing for the 90 participants who were given that opportunity (see Section 3.2.2). The second round results are broken down in the column labeled *2nd round guess improvement*. As 40 participants did not have this opportunity, the *guessed by partner* columns under-represent the ability of partners to guess successfully.

What is your pet's name? is asked by both AOL and Yahoo!, but is only counted once in the totals in the bottom row. Thus, these totals are less than the sum of all rows.

For all answers to all questions, 13% were statistically guessable. An attacker with a larger sample of answers than we used might be able to do even better. Given that almost all participants were from the same metropolitan area, it's not surprising that their favorite sports teams can be guessed more than half the time and many had the same favorite towns. However, other questions had popular answers that seem unlikely to change much within the United States. Favorite pastimes (e.g. travel, reading) are fairly geographically universal activities, and both childhood heros (e.g. Superman) and historical persons (e.g. Jesus Christ) are often drawn from a culture that is growing ever more globalized.

5.5. Security against guessing by acquaintance

For all participants who answered a question, the fourth data column of Table 4 (labeled *guessed by partner*) shows the number and percentage of those participants' partners were able to guess their answers. These figures include guesses made in the second guessing round, at the end of the laboratory session, though the 40 participants in the first cohort did not have this guessing opportunity. The fourth, fifth, and sixth columns show these figures broken down based on how participants responded to the question *would you trust your partner with your Hotmail password?* to which they might answer *no*, *yes*, or *under some circumstances*. Highlighted are percentages above 20% in the *no* column.

Google's questions performed the best by this metric. Nobody guessed the answer to their partner's ID card answers and the overall guess rate was just 4%. Microsoft's questions came in a very distant second at 16% and AOL and Yahoo! trailed at 26% and 27%, respectively.

Participants who were not trusted with their partners' Hotmail password, or who were trusted only *under some circumstances*, had roughly equal success in guessing their partners' answers. Both groups were able to guess the answers to roughly 17% of their partners' questions. In contrast, those who were trusted by their partners were able to guess their partners' answers 28% of the time. We ran a t-test to measure the effect of this trust question on the percentage of answers guessed by participants' partners. The difference between partners who participants would trust with their password and those who they would never trust did not meet the threshold of significance, $t(88) = -1.750, p = .0860$. The difference between partners who were fully trusted and all others (both *no* and *under some circumstances* to the trust question) was strongly significant, $t(130) = -3.096, p = .002$.

Questions with answers that participants found easiest to recall appeared to be those that their partners found easiest to guess. A non-parametric Kendall τ test, examining the correlation between the fraction of answers recalled for each question and the fraction guessed by participants' partners, indicates a strong correlation, $\tau(56) = .496, p < 0.001$.

There are numerous reasons to believe these numbers may underestimate the ability of participants' partners to guess their answers. A third of the questions received no guesses and when participants did guess, most used only one of their their five allotted guesses. Participants would likely have done better if they had received feedback on which guesses were correct. For example, in Table 7 (in the Appendix) we see that half of the participants whose partner was their spouse were unable to guess how their partner answered the question *where did you meet your spouse?*.

Furthermore, the first cohort of 40 participants were not asked to perform online research and not given a second guessing round at the end of the laboratory session. The rightmost column in Table 4 shows the influence of the second round on the results of those participants who had been given this second opportunity. Not surprisingly, questions like *high school mascot* become easier to guess with online research.

5.6. The security of user-written questions

A total of 127 of our 130 participants responded to our request to write their own question and provide the answer.

User-written questions are harder for us to analyze, and possibly harder to attack, in as automated a manner as site-written questions. This does not mean they are immune to attack. In our study, seven participants (6%) chose questions that matched, or were significantly similar to site-written questions and are likely to have similar answer distributions. These questions are listed in Table 5.1.

Through manual analysis, we identified concrete problems with 63 of the 120 user-written question/answer pairs that remained, which constituted half of all user-written question/answer pairs. We broke these into two subcategories, each of which contained roughly a quarter of all question/answer pairs.

The first subcategory, in Table 5.2, contains 31 question/answer pairs (24% of all pairs) vulnerable to attacks that require no personal knowledge beyond the geographic location of the account holder. Using the techniques described in the table, nineteen of these

Table 5. If we allowed you to write your own personal question for your own use, what would it be?

1. Questions similar to those already used by webmail services (7 of 127, 6%)

Question written by participant	*Similar question asked by webmail service*
What is your first job	Where was your first job? (AOL)
What is my favorite thing to do?	What is your favorite pastime? (Yahoo!)
my childhood best friend	Best childhood friend (Microsoft)
First Car	What make was your first car or bike? (Yahoo!)
What is the name of your first pet as an adult?	What is your pets name? (AOL & Yahoo!)
First pet's first name	What is your pets name? (AOL & Yahoo!)
What is your favorite pet's name?	What is your pets name? (AOL & Yahoo!)

2. Vulnerable with no personal knowledge other than geographic region (31 of 127, 24%)

	answer space: size (categorized into 5,10,25) & how obtained	
i. Answer can be found via simple web search (2, 2%)		
The story of a Number	5	"the story of a number" (answer is top hit)
What's your favorite cookie at Panera Bakery?	5	"panera cookies" (answer in five cookie names listed on menu)
ii. Answer space ≤ 5 (11, 8%), ≤ 10 (15, 12%) & ≤ 25 (18, 14%)		
Water or Pop?	5	"water", "pop"
How many children do I have?	5	Count up from 0, "zero"
Favorite kink	5	Four examples in wikipedia definition
when did i graduate college?	5	Years backwards from 2008
What color are your eyes?	5	Four most common color names ("brown", "hazel", "green", "blue")
What is the color of your eyes	5	Four most common color names ("brown", "hazel", "green", "blue")
what color was your triumph	5	Three primary colors ("red", "green", "blue")
Who should our next President be?	5	Presidential candidates (spring 2008)
What is my blood type	5	Four primary blood types ("o", "a", "b", "ab")
Where do I want to be living in 10 years?	5	Local city names (by size)
Where do you live?	5	Local city names (by size)
How tall am I?	10	Count up from 5'0" and "five foot zero"
how tall are you	10	Count up from 5'0" and "five foot zero"
What inseam do you wear?	10	Count up from 26, "26 inches"
number of times i got stitches?	10	Count up from 0, "zero"
What is your favorite number?	25	Count up from 0, "zero"
What was the year you graduated high school?	25	Years backwards from 2008
how many words I can type in on minute	25	Numbers around average typing speed in US
iii. Answer high on easily searchable popularity lists, top 5 (6, 5%), top 25 (11, 7%)		
What is your have soda? [sic]	5	Best selling sodas in US
Favorite Food	5	Most popular foods in US
what sports team would you love to see lose	5	Most popular or top grossing sports teams
Which sports team do you love to hate?	5	Most popular or top grossing sports teams
Favorite TV show	5	Highest rated TV shows
First car	5	Top selling auto makers
Which is my favorite holiday?	25	Most popular holidays
Favorite Beer?	25	Top beer brands (US)
favorite beer	25	Top beer brands (US)
Who is your favorite actor?	25	Top grossing actors of all time
Best video game ever created?	25	Top selling games of all time

3. Vulnerable to coworkers, clients, or family members (32 of 127, 25%)

i. Vulnerable to family members (29, 23%)	**ii. Vulnerable to coworkers/clients (3, 2%)**
question	*question*
mother's maiden name? (10 occurrences)	Who is your current boss?
mothers middle name (4 occurrences)	What is my line of work
father's first name (2 occurrences)	What do you keep on your desk at all times?
Daughter's Middle Name	
favorite relative's name	
First child's middle name	
first initial of your sisters names from oldest to youngest	
mothers name?	
name of my wife	
place of child birth	
place where you were married	
significant other's middle name?	
StepFathers middle name	
What is [child's name]'s favorite toy?	
when were you married	
your birthdate	
your children's godparents	

pairs (15% of all pairs) would be guessed within 5 attempts. Two of these pairs had answers that could be easily identified via a simple web search. Another eleven drew from a small answer space, such as eye colors, with the answer falling within that space. Another six answers were within the top five results of easily searchable online popularity lists. For example, one participant's question was *Favorite TV show* and among the top five rated shows according to the first popularity list we searched—the Nielson television ratings. Note that our statistics and the list of question/answer pairs in Table 5.2.iii includes only those pairs with popular answers; we exclude pairs with the same or similar questions for which participants responded with less popular answers. For example, two pairs we excluded included the question *favorite food* but had answers that were not at the top of popularity lists.

The second subcategory, in Table 5.3, contains 32 question/answer pairs (25% of all pairs) vulnerable to attack by family members or others the participants knew. Fourteen participants asked for either their mother's maiden name or their mother's middle name. A total of 29 questions (23% of all user-written questions) were categorized as vulnerable to family members. Another three questions were clearly vulnerable to coworkers and clients, though many from the family category (especially *name of my wife*) would also be vulnerable.

Finally, a few participants may have not understood that while the answers they wrote would not be public, the questions would be. One proposed question, *my sobriety date*, would be a poor choice if the participant considered his history of alcoholism to be private. Another user-written question, *what is my favorite kink?*, might also have conveyed more information than intended by the participant who wrote it.

Advocates of user-written questions might argue that these questions would have worked better if we had only taught participants how to choose a strong question; we are skeptical. Users would have to take the time to read or view the instructions, learn and understand the different types of threats to their answers, generate a sufficient number of candidate question/answer pairs to come across one that is both memorable and secure against all threats, and reject all pairs that failed to meet these complex criteria.

6. Discussion

Our results do not give us confidence that today's personal questions make adequate authentication secrets. Those that are hard to guess are less likely to be chosen by users in the first place, and when chosen they are less likely to be remembered.

While the most well publicized attacks on personal questions have been targeted at individuals, our results show that large scale attacks are also possible. Blackhats already have mailing lists containing large lists of user accounts hosted by webmail services. Our results show that a significant fraction of these accounts could be compromised simply by providing the most popular answers to users' personal authentication questions.

Furthermore, there are a number of other threats against personal questions that we did not address in our study. The two questions that went unguessed during the study asked for ID numbers: frequent flyer numbers and library card numbers. The accounts of users who choose the former, and who use their webmail account to communicate with airlines and travel agencies, can be compromised by anyone with access to these firms' databases. Airlines, travel agencies, and libraries may not guard these ID numbers with the same vigilance that a user would expect his or her password to be guarded with. The answers to these questions may also be easier for an attacker to obtain than they were for our participants to guess in the laboratory. For example, an attacker might offer the owner of a target account a prize if she can prove she traveled in the last month—"just send an itinerary as proof of travel".

6.1. Improving questions

Many shared secret authentication schemes, including those that use personal questions, limit the user to a fixed threshold of responses (answers). One way to make secret questions more secure and reliable would be to dynamically adjust the threshold based on the types of responses received. In other words, certain responses are penalized (move a user closer to the threshold) than others.

To reduce vulnerability to statistical guessing attacks, responses could be penalized in proportion to their popularity. This could limit attackers to two or three popular answers. The size of the penalty would depend on the likelihood that a legitimate user would respond with multiple popular answers before guessing the correct one. Table 9 illustrates that of the 900 answers eventually recalled by participants in our follow-up study, 44 (5%) were preceded by a response that was both incorrect and that matched one of the five most popular answers for that question. Only one of the 900 correct answers (0.1%) was preceded by two incorrect but popular answers.

Users who are trying to recall a correct answer may provide answers that are similar to each other. For example, a user who had no trouble remembering where her mother was born might still walk through numerous possible answers: 'Coney Island', 'coney island', 'Brooklyn', 'Brooklyn, NY', 'Brooklyn, New York', 'New York', 'New York, NY', and so on. A user should not be penalized for a response that is identical to a previous response for the purposes of authentication, and should be penalized less when responses are similar to each other or to the correct answer. For example, 'Coney Island' and 'coney island' are lexographically identical with the exception of capitalization, and so the latter response should not be penalized. The responses 'Brooklyn' and 'Brooklyn, NY' are lexographically similar—they share a common prefix. Other responses, such as 'Coney Island' and 'Brooklyn', are semantically similar as one might infer using a geographic database. When similarity is not as easy to obtain as in this example, previous users' responses might be mined to reveal commonly confused answers.

Another way to reduce vulnerability to statistical guessing attacks is to reduce the proportion of popular answers. We propose eliminating questions that are currently statistically guessable more than 10% of the time. For the remaining questions, we propose flagging and rejecting answers that exceed a certain threshold of popularity (e.g. 1%). Users would be asked to choose another question or a more specific answer.

Unpopular answers may also be harder for acquaintances to guess. For all 379 answers deemed statistically guessable, 168 (44%) were guessed by participants' partners. In contrast, only 460 of the 2491 answers that were not deemed statistically guessable (18%) were guessed by participants' partners. A Fisher's exact test shows the difference to be statistically significant, $p < 0.0001$.

Guiding users away from popular answers may also increase the likelihood that they will forget them. We suggest occasionally inserting a query for each user's answer after login has completed. We would encourage those who have trouble recalling their answers to choose a new answer (or an entirely new question). The first such query should occur shortly after a question is configured – perhaps a few days – to ensure the answer was encoded to long-term memory. Additional queries could be separated by much longer periods (e.g. six months) and ensure the answers had not changed.

Some might be concerned that an authentication system that alerts users when they have chosen a popular answer could be used by attackers as an oracle to identify these popular answers—it could. However, with little effort an attacker can collect answers from the public and derive more accurate popularity statistics than could be obtained from the authentication system.

Some websites' backup authentication systems allow users to configure hints that will help them recall the correct answer in the future. While we did not examine this practice, our findings on user-written questions leave us concerned that users might be unable to sufficiently tune hints to remind them of their answers without revealing these answers to others.

Other websites' backup authentication systems require users to configure multiple questions and answer a subset to authenticate. Designers of such systems must decide whether to reveal which answers a user got correct if he or she fails to provide a sufficient number of correct answers. This is likely to be a common case, as we found 24% of answers that were eventually recalled by our participants were not correctly recalled in the first guess. (For more detailed statistics, see Table 8 in the Appendix.) If users were asked all questions at once and not told which questions they answered correctly and which they had not, many users who would have been able to answer a sufficient number of questions asked individually would no longer be able to do so. On the other hand, if incorrect answers were individually identified, adversaries could determine how close they were to a sufficient number of answers and which they needed to research further. Furthermore, using multiple questions might lull users into believing that it is safe to reveal individual answers, thinking that the remaining questions are likely a sufficient defense.

One approach to backup authentication using multiple questions, proposed by Jakobsson *et al.* [4], relies on preference-based questions, similar to those on online dating websites, with answers rated on a scale. However, this approach requires both a large number of questions to be configured and a large number of responses during authentication.

6.2. Alternative backup authenticators

One barrier to the deployment of new, and potentially better, backup authentication options is that the comparative risks are unknown. We hope that by quantifying the risks of personal questions, we will help to catalyze the development of quantitatively-superior alternatives.

One current alternative, authentication via a code sent to an alternate email address, is often not viable for users' primary email accounts. Even when users have alternate addresses they can provide, these addresses may expire when users change their ISP,

school, job, or other affiliation. Simultaneous credential loss could occur if a user stored her password on a work computer, used her work email address as her backup authenticator, and then lost her job.

Mobile phones are already in use as a second authentication factor by some banks [3], which send authentication codes to users in SMS messages. Authentication using mobile phones is attractive because of phones' ubiquity. However, phones are also frequently shared, lost, and stolen. The security of SMS message transmission is also a concern.

Many users protect against memory loss by writing passwords down. Rather than admonish them for this practice, a backup authentication system could instead offer to print a list of single-use account-recovery codes and encourage users to store them in a locked filing cabinet, safe, or safe-deposit boxes. As with written passwords, a printed list might not be available when the user was away from the location(s) at which it was stored. Furthermore, simultaneous credential loss could occur if a user stored her password in her browser, stored her authentication list in a safe near the computer, and then lost both in a natural disaster.

In previous papers, Brainard *et al.* [1] and we [12] have proposed and tested systems in which user-selected trustees vouch for the identity of the user. While early reliability and security results from our work show promise, communicating with one or more trustees requires far more work than typing a simple answer to a question. For many users, the consequences of having an account lost or compromised may not be significant enough to justify the extra effort.

7. Conclusion

Backup authentication mechanisms should reliably enable account holders to regain access to accounts for which they have forgotten their passwords, and do so without significantly increasing the risk that the account can be compromised.

The secret questions employed by the top four webmail services are not sufficiently reliable authenticators. Even for the webmail service with the most memorable set of questions (Yahoo!), participants forgot an average of 16% of the answers to those questions within six months.

The security of personal questions appears significantly weaker than passwords. Acquaintances with whom participants reported being unwilling to share their Hotmail passwords were able to guess 17% of answers. For our geographically-homogenous sample, 13% of answers could be guessed by iterating through the five most popular answers of other users. User-written questions were no better: roughly half were vulnerable to guessing by either acquaintances or those who had never met the account holder.

Whatever options users are given for backup authentication, all have risks and users have the right to know about them. We hope this work helps users to choose whether and how to answer backup authentication questions. We also hope that by quantifying the bar over which new backup authentication mechanisms must pass, we will inspire the creation, measurement, and deployment of new alternatives to 'secret' questions.

Acknowledgments

This paper was inspired by the griping of Jon Howell. We are indebted to Will Ip, Maritza Johnson, and Arry Shin for their assistance in running our study. We are also grateful for the valuable feedback on earlier drafts provided by Robert W. Reeder and the anonymous reviewers.

Epilog

On November 12, 2008, we contacted AOL, Google, and Yahoo! to provide them with a draft of this paper and share our intent to publish at this symposium. We asked to be notified by the end of 2008 if they had concerns that might warrant the delay of publication, so as to provide ample time to discuss these concerns with them and, if necessary, withdraw the paper. AOL and Google sent email explicitly consenting to publication in advance of the deadline. Yahoo! made no request to delay publication. We learned in February 2009 that Yahoo! had replaced all nine of the personal authentication questions that its users may choose from when signing up for a new account.

References

[1] J. Brainard, A. Juels, R. L. Rivest, M. Szydlo, and M. Yung. Fourth-factor authentication: somebody you know. In *CCS '06: Proceedings of the 13th ACM Conference on Computer and Communications Security*, pages 168–178, New York, NY, USA, 2006. ACM.

[2] T. Bridis. Hacker impersonated Palin, stole e-mail password, Sept. 18, 2008. Associated Press.

[3] CommonwealthBank. NetBank NetCode SMS, 2008. http://www.commbank.com.au/netbank/netcodesms/.

[4] M. Jakobsson, E. Stolterman, S. Wetzel, and L. Yang. Love and authentication. In *CHI '08: Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems*, pages 197–200, New York, NY, USA, 2008. ACM.

[5] M. Just. Designing authentication systems with challenge questions. In L. F. Cranor and S. Garfinkel, editors, *Security and Usability: Designing Secure Systems that People Can Use*, pages 143–155, Sebastopol, CA, 2005. O'Reilly Media, Inc.

[6] G. Keizer. Yahoo, Hotmail, Gmail all vulnerable to Palin-style password-reset hack. *Computerworld*, Sept. 19, 2008. http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9115187.

[7] J. Kremer. Happy 10th birthday, Yahoo! Mail, Oct. 2007. http://ycorpblog.com/2007/10/08/happy-10th-birthday-yahoo-mail/.

[8] H. P. Ltd. Top 20 websites, 2008. http://www.hitwise.com/datacenter/rankings.php.

[9] Microsoft Corporation. Windows live hotmail fact sheet, May 2007. http://www.microsoft.com/presspass/newsroom/msn/factsheet/hotmail.mspx.

[10] J. Podd, J. Bunnell, and R. Henderson. Cost-effective computer security: Cognitive and associative passwords. In *OZCHI '96: Proceedings of the 6th Australian Conference on Computer-Human Interaction (OZCHI '96)*, page 304, Washington, DC, USA, 1996. IEEE Computer Society.

[11] A. Rabkin. Personal knowledge questions for fallback authentication: security questions in the era of facebook. In *SOUPS '08: Proceedings of the 4th Symposium on Usable Privacy and Security*, pages 13–23, New York, NY, USA, 2008. ACM.

[12] S. Schechter, S. Egelman, and R. W. Reeder. It's not what you know, but who you know: A social approach to last-resort authentication. In *CHI '09: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, Boston, MA, 2009. ACM.

[13] S. E. Schechter, R. Dhamija, A. Ozment, and I. Fischer. The emperor's new security indicators: An evaluation of website authentication and the effect of role playing on usability studies. In *Proceedings of the 2007 IEEE Symposium on Security and Privacy*, pages 51–65, Washington, DC, USA, May 20–23 2007. IEEE Computer Society.

[14] R. Stross. What would you do if you logged onto your e-mail and received an unfamiliar message: 'user name and password do not match'? *The New York Times*, Oct. 4, 2008. http://www.nytimes.com/2008/10/05/business/05digi.html.

[15] B. Sullivan. 'forgot your password?' may be weakest link. *MSNBC Red Tape Chronicles*, Aug. 26, 2008. http://redtape.msnbc.com/2008/08/almost-everyone.html.

[16] M. Toomim, X. Zhang, J. Fogarty, and J. A. Landay. Access control by testing for shared knowledge. In *CHI '08: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, Florence, Italy, 2008. ACM.

[17] M. Zviran and W. J. Haga. User authentication by cognitive passwords: an empirical assessment. In *JCIT: Proceedings of the Fifth Jerusalem Conference on Information technology*, pages 137–144, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press.

Appendix A. The missing question

One of the personal authentication questions used by Microsoft is *name of first pet*. Due to a clerical error, our original survey instead asked participants the question *your first pet*. The removal of "name of" had an important effect on the answers: many were simply *dog* or *cat*. Further confounding the problem, we asked the correct question (*name of first pet*) during the longitudinal recall study. Thus, we had no choice but to exclude these results from our findings. We audited all other questions and found no other errors of this type.

The excluded question was similar to a question shared by AOL and Yahoo: *what is your pet's name?*. One important difference is that Microsoft's question asks specifically about a first pet, which may be less well known to acquaintances than one's current pet. If the results for Microsoft's question were equivalent to those for *what is your pet's name*, the aggregate vulnerability of all Microsoft questions to guessing by partners would have increased. However, aggregate statistics for both participant recall (memorability) and resilience to statistical guessing would have improved.

Table 6. Guesses broken down by partner relationship

	Spouse *18 ppts*	**Relative** *23 ppts*	**Fiance/SO** *4 ppts*	**Friend** *51 ppts*	**Coworker** *32 ppts*	**Other** *2 ppts*
AOL						
What is your pet's name?	13/14 (93%)	8/16 (50%)	2/4 (50%)	17/36 (47%)	5/22 (23%)	0/1 (0%)
Where were you born?	13/18 (72%)	15/23 (65%)	2/4 (50%)	20/51 (39%)	10/31 (32%)	0/2 (0%)
What is your favorite restaurant?	5/17 (29%)	2/19 (11%)	0/3 (0%)	12/48 (25%)	1/29 (3%)	0/1 (0%)
What is the name of your school?	3/14 (21%)	3/17 (18%)	0/2 (0%)	14/39 (36%)	7/22 (32%)	0/2 (0%)
Who is your favorite singer?	3/15 (20%)	3/15 (20%)	0/2 (0%)	5/42 (12%)	1/27 (4%)	0/0
What is your favorite town?	6/17 (35%)	5/21 (24%)	0/2 (0%)	13/47 (28%)	9/28 (32%)	0/0
What is your favorite song?	1/14 (7%)	0/13 (0%)	0/2 (0%)	1/40 (3%)	2/24 (8%)	0/1 (0%)
What is your favorite film?	6/15 (40%)	0/18 (0%)	1/3 (33%)	9/50 (18%)	2/28 (7%)	0/0
What is your favorite book?	2/13 (15%)	3/17 (18%)	0/3 (0%)	5/41 (12%)	2/25 (8%)	0/1 (0%)
Where was your first job?	5/18 (28%)	4/20 (20%)	1/4 (25%)	13/51 (25%)	3/31 (10%)	0/1 (0%)
Where did you grow up?	9/18 (50%)	11/22 (50%)	0/4 (0%)	26/50 (52%)	15/31 (48%)	0/2 (0%)
Total	66/173 (38%)	54/201 (27%)	6/33 (18%)	135/495 (27%)	57/298 (19%)	0/11 (0%)
Google						
What is your primary frequent flyer number?	0/4 (0%)	0/6 (0%)	0/2 (0%)	0/8 (0%)	0/8 (0%)	0/1 (0%)
What is your library card number?	0/4 (0%)	0/4 (0%)	0/3 (0%)	0/19 (0%)	0/9 (0%)	0/0
What was your first phone number?	3/13 (23%)	3/13 (23%)	0/2 (0%)	2/38 (5%)	1/27 (4%)	0/0
What was your first teacher's name?	0/13 (0%)	1/13 (8%)	0/2 (0%)	0/38 (0%)	0/27 (0%)	0/0
Total	3/34 (9%)	4/36 (11%)	0/9 (0%)	2/103 (2%)	1/71 (1%)	0/1 (0%)
Microsoft						
Mother's birthplace	8/16 (50%)	8/20 (40%)	0/4 (0%)	12/49 (24%)	6/30 (20%)	0/2 (0%)
Best childhood friend	9/17 (53%)	5/20 (25%)	0/3 (0%)	8/49 (16%)	1/29 (3%)	0/2 (0%)
Favorite teacher	0/16 (0%)	1/16 (6%)	0/2 (0%)	4/43 (9%)	0/28 (0%)	0/0
Favorite historical person	4/17 (24%)	1/16 (6%)	0/2 (0%)	6/47 (13%)	2/24 (8%)	0/0
Grandfather's occupation	3/15 (20%)	3/15 (20%)	1/3 (33%)	3/39 (8%)	2/25 (8%)	0/2 (0%)
Total	24/81 (30%)	18/87 (21%)	1/14 (7%)	33/227 (15%)	11/136 (8%)	0/6 (0%)
Yahoo!						
Where did you meet your spouse?	9/18 (50%)	7/14 (50%)	2/3 (67%)	4/29 (14%)	1/15 (7%)	0/1 (0%)
What was the name of your first school?	4/18 (22%)	8/17 (47%)	0/2 (0%)	8/49 (16%)	3/30 (10%)	0/0
Who was your childhood hero?	1/13 (8%)	1/19 (5%)	0/1 (0%)	3/37 (8%)	5/27 (19%)	0/0
What is your favorite pastime?	6/18 (33%)	6/22 (27%)	2/4 (50%)	10/45 (22%)	4/28 (14%)	0/1 (0%)
What is your favorite sports team?	11/16 (69%)	9/19 (47%)	0/0	17/40 (43%)	10/24 (42%)	0/1 (0%)
What is your father's middle name?	8/16 (50%)	6/16 (38%)	0/4 (0%)	3/43 (7%)	0/27 (0%)	0/2 (0%)
What was your high school mascot?	7/14 (50%)	8/20 (40%)	0/2 (0%)	14/44 (32%)	5/29 (17%)	0/0
What make was your first car or bike?	5/18 (28%)	7/20 (35%)	0/4 (0%)	13/51 (25%)	5/31 (16%)	1/2 (50%)
What is your pet's name?	13/14 (93%)	8/16 (50%)	2/4 (50%)	17/36 (47%)	5/22 (23%)	0/1 (0%)
Total	64/145 (44%)	60/163 (37%)	6/24 (25%)	89/374 (24%)	38/233 (16%)	1/8 (13%)
Total for all webmail sites	144/419 (34%)	128/471 (27%)	11/76 (14%)	242/1163 (21%)	102/716 (14%)	1/25 (4%)

Table 7. Guesses broken down by how long partners knew each other

	< 6 months *6 ppts*	**6 months–1 year** *11 ppts*	**1–4 years** *30 ppts*	**> 4 years** *83 ppts*
AOL				
What is your pet's name?	0/3 (0%)	0/4 (0%)	7/23 (30%)	38/63 (60%)
Where were you born?	3/5 (60%)	3/11 (27%)	11/30 (37%)	43/83 (52%)
What is your favorite restaurant?	0/5 (0%)	1/9 (11%)	1/28 (4%)	18/75 (24%)
What is the name of your school?	2/3 (67%)	3/9 (33%)	7/21 (33%)	15/63 (24%)
Who is your favorite singer?	0/5 (0%)	0/9 (0%)	0/24 (0%)	12/63 (19%)
What is your favorite town?	3/5 (60%)	3/9 (33%)	5/26 (19%)	22/75 (29%)
What is your favorite song?	0/3 (0%)	0/9 (0%)	1/24 (4%)	3/58 (5%)
What is your favorite film?	0/5 (0%)	2/10 (20%)	3/30 (10%)	13/69 (19%)
What is your favorite book?	0/5 (0%)	1/8 (13%)	3/25 (12%)	8/62 (13%)
Where was your first job?	1/5 (20%)	2/11 (18%)	3/30 (10%)	20/79 (25%)
Where did you grow up?	2/5 (40%)	3/11 (27%)	13/30 (43%)	43/81 (53%)
Total	11/49 (22%)	18/100 (18%)	54/291 (19%)	235/771 (30%)
Google				
What is your primary frequent flyer number?	0/0	0/1 (0%)	0/8 (0%)	0/20 (0%)
What is your library card number?	0/1 (0%)	0/3 (0%)	0/8 (0%)	0/27 (0%)
What was your first phone number?	0/4 (0%)	0/8 (0%)	1/24 (4%)	8/57 (14%)
What was your first teacher's name?	0/4 (0%)	0/7 (0%)	0/27 (0%)	1/55 (2%)
Total	0/9 (0%)	0/19 (0%)	1/67 (1%)	9/159 (6%)
Microsoft				
Mother's birthplace	1/5 (20%)	1/10 (10%)	5/30 (17%)	27/76 (36%)
Best childhood friend	0/4 (0%)	0/10 (0%)	2/29 (7%)	21/77 (27%)
Favorite teacher	0/5 (0%)	0/10 (0%)	1/28 (4%)	4/62 (6%)
Favorite historical person	0/3 (0%)	2/9 (22%)	1/24 (4%)	10/70 (14%)
Grandfather's occupation	1/2 (50%)	0/7 (0%)	2/26 (8%)	9/64 (14%)
Total	2/19 (11%)	3/46 (7%)	11/137 (8%)	71/349 (20%)
Yahoo!				
Where did you meet your spouse?	1/1 (100%)	0/5 (0%)	2/14 (14%)	20/60 (33%)
What was the name of your first school?	1/5 (20%)	0/11 (0%)	2/28 (7%)	20/72 (28%)
Who was your childhood hero?	1/4 (25%)	1/9 (11%)	3/24 (13%)	5/60 (8%)
What is your favorite pastime?	0/3 (0%)	2/9 (22%)	5/27 (19%)	21/79 (27%)
What is your favorite sports team?	3/4 (75%)	1/7 (14%)	8/21 (38%)	35/68 (51%)
What is your father's middle name?	0/4 (0%)	0/9 (0%)	0/26 (0%)	17/69 (25%)
What was your high school mascot?	2/5 (40%)	2/9 (22%)	2/28 (7%)	28/67 (42%)
What make was your first car or bike?	2/5 (40%)	2/11 (18%)	3/30 (10%)	24/80 (30%)
What is your pet's name?	0/3 (0%)	0/4 (0%)	7/23 (30%)	38/63 (60%)
Total	10/34 (29%)	8/74 (11%)	32/221 (14%)	208/618 (34%)
Total for all webmail sites	23/108 (21%)	29/235 (12%)	91/693 (13%)	485/1834 (26%)

Table 8. Guesses required to recall those answers that could eventually be recalled

	Guess Number									
	1	2	3	4	5	6	7	8	9	10
AOL										
What is your pet's name?	89%	97%	97%	100%	100%	100%	100%	100%	100%	100%
Where were you born?	85%	91%	94%	100%	100%	100%	100%	100%	100%	100%
What is your favorite restaurant?	74%	91%	91%	97%	97%	100%	100%	100%	100%	100%
What is the name of your school?	67%	77%	87%	97%	97%	97%	97%	100%	100%	100%
Who is your favorite singer?	72%	76%	100%	100%	100%	100%	100%	100%	100%	100%
What is your favorite town?	71%	94%	94%	100%	100%	100%	100%	100%	100%	100%
What is your favorite song?	62%	86%	86%	90%	90%	95%	100%	100%	100%	100%
What is your favorite film?	61%	75%	82%	96%	96%	100%	100%	100%	100%	100%
What is your favorite book?	78%	93%	96%	100%	100%	100%	100%	100%	100%	100%
Where was your first job?	59%	77%	87%	90%	97%	97%	97%	100%	100%	100%
Where did you grow up?	76%	93%	96%	100%	100%	100%	100%	100%	100%	100%
Total	73%	87%	92%	98%	98%	99%	99%	100%	100%	100%
Google										
What is your primary frequent flyer number?	60%	100%	100%	100%	100%	100%	100%	100%	100%	100%
What is your library card number?	88%	100%	100%	100%	100%	100%	100%	100%	100%	100%
What was your first phone number?	87%	97%	97%	100%	100%	100%	100%	100%	100%	100%
What was your first teacher's name?	75%	92%	96%	100%	100%	100%	100%	100%	100%	100%
Total	81%	96%	97%	100%	100%	100%	100%	100%	100%	100%
Microsoft										
Mother's birthplace	77%	90%	95%	95%	97%	100%	100%	100%	100%	100%
Best childhood friend	79%	92%	92%	92%	92%	92%	95%	97%	97%	97%
Favorite teacher	61%	87%	90%	94%	97%	97%	97%	97%	97%	100%
Favorite historical person	68%	80%	92%	96%	100%	100%	100%	100%	100%	100%
Grandfather's occupation	70%	78%	81%	89%	93%	93%	96%	100%	100%	100%
Total	72%	86%	91%	93%	96%	96%	98%	99%	99%	99%
Yahoo!										
Where did you meet your spouse?	60%	80%	87%	90%	93%	93%	93%	97%	100%	100%
What was the name of your first school?	73%	85%	90%	95%	95%	95%	98%	98%	100%	100%
Who was your childhood hero?	50%	78%	89%	94%	94%	94%	94%	94%	94%	94%
What is your favorite pastime?	63%	70%	83%	90%	97%	100%	100%	100%	100%	100%
What is your favorite sports team?	71%	94%	97%	100%	100%	100%	100%	100%	100%	100%
What is your father's middle name?	95%	100%	100%	100%	100%	100%	100%	100%	100%	100%
What was your high school mascot?	95%	100%	100%	100%	100%	100%	100%	100%	100%	100%
What make was your first car or bike?	66%	91%	98%	98%	98%	98%	100%	100%	100%	100%
What is your pet's name?	89%	97%	97%	100%	100%	100%	100%	100%	100%	100%
Total	76%	90%	94%	97%	98%	98%	99%	99%	100%	100%
All questions	74%	89%	93%	97%	98%	98%	99%	99%	100%	100%

Each column i represents the number of answers guessed within the first i tries as a percentage of the number of answers that could be recalled given an unlimited number of attempts. (In most of this paper, participants are said to have forgotten their answer if they fail in the first five attempts.)

Table 9. Statistical guessing and popular answers

	answers among five most popular		answers deemed statistically guessable		# of incorrect but popular (among top five) responses before correct answer recalled 0		1		2	
AOL										
What is your pet's name?	10/93	(11%)	0/93	(0%)	37/37	(100%)	0/37	(0%)	0/37	(0%)
Where were you born?	27/129	(21%)	19/129	**(15%)**	43/45	(96%)	2/45	(4%)	0/45	(0%)
What is your favorite restaurant?	19/117	(16%)	7/117	(6%)	32/34	(94%)	2/34	(6%)	0/34	(0%)
What is the name of your school?	22/96	(23%)	22/96	**(23%)**	30/30	(100%)	0/30	(0%)	0/30	(0%)
Who is your favorite singer?	11/101	(11%)	1/101	(1%)	28/28	(100%)	0/28	(0%)	0/28	(0%)
What is your favorite town?	37/115	(32%)	34/115	**(30%)**	31/35	(89%)	4/35	(11%)	0/35	(0%)
What is your favorite song?	3/94	(3%)	3/94	(3%)	21/21	(100%)	0/21	(0%)	0/21	(0%)
What is your favorite film?	18/114	(16%)	11/114	(10%)	26/28	(93%)	2/28	(7%)	0/28	(0%)
What is your favorite book?	21/100	(21%)	19/100	**(19%)**	27/27	(100%)	0/27	(0%)	0/27	(0%)
Where was your first job?	16/125	(13%)	10/125	(8%)	34/37	(92%)	3/37	(8%)	0/37	(0%)
Where did you grow up?	27/127	(21%)	22/127	**(17%)**	42/43	(98%)	1/43	(2%)	0/43	(0%)
Total	211/1211	(17%)	148/1211	(12%)	351/365	(96%)	14/365	(4%)	0/365	(0%)
Google										
What is your primary frequent flyer number?	0/29	(0%)	0/29	(0%)	5/5	(100%)	0/5	(0%)	0/5	(0%)
What is your library card number?	0/39	(0%)	0/39	(0%)	7/7	(100%)	0/7	(0%)	0/7	(0%)
What was your first phone number?	0/93	(0%)	0/93	(0%)	30/30	(100%)	0/30	(0%)	0/30	(0%)
What was your first teacher's name?	6/93	(6%)	6/93	(6%)	24/24	(100%)	0/24	(0%)	0/24	(0%)
Total	6/254	(2%)	6/254	(2%)	66/66	(100%)	0/66	(0%)	0/66	(0%)
Microsoft										
Mother's birthplace	19/121	(16%)	12/121	(10%)	34/38	(89%)	4/38	(11%)	0/38	(0%)
Best childhood friend	10/120	(8%)	1/120	(1%)	39/39	(100%)	0/39	(0%)	0/39	(0%)
Favorite teacher	0/105	(0%)	0/105	(0%)	31/31	(100%)	0/31	(0%)	0/31	(0%)
Favorite historical person	30/106	(28%)	27/106	**(25%)**	21/24	(88%)	3/24	(13%)	0/24	(0%)
Grandfather's occupation	20/99	(20%)	13/99	**(13%)**	25/27	(93%)	2/27	(7%)	0/27	(0%)
Total	79/551	(14%)	53/551	(10%)	150/159	(94%)	9/159	(6%)	0/159	(0%)
Yahoo!										
Where did you meet your spouse?	19/80	(24%)	10/80	**(13%)**	25/29	(86%)	4/29	(14%)	0/29	(0%)
What was the name of your first school?	9/116	(8%)	1/116	(1%)	41/41	(100%)	0/41	(0%)	0/41	(0%)
Who was your childhood hero?	34/97	(35%)	27/97	**(28%)**	15/18	(83%)	3/18	(17%)	0/18	(0%)
What is your favorite pastime?	32/118	(27%)	23/118	**(19%)**	24/27	(89%)	3/27	(11%)	0/27	(0%)
What is your favorite sports team?	59/100	(59%)	57/100	**(57%)**	30/35	(86%)	5/35	(14%)	0/35	(0%)
What is your father's middle name?	17/108	(16%)	5/108	(5%)	40/40	(100%)	0/40	(0%)	0/40	(0%)
What was your high school mascot?	21/109	(19%)	18/109	**(17%)**	39/39	(100%)	0/39	(0%)	0/39	(0%)
What make was your first car or bike?	36/126	(29%)	31/126	**(25%)**	37/44	(84%)	6/44	(14%)	1/44	(2%)
What is your pet's name?	10/93	(11%)	0/93	(0%)	37/37	(100%)	0/37	(0%)	0/37	(0%)
Total	237/947	(25%)	172/947	(18%)	288/310	(93%)	21/310	(7%)	1/310	(0.3%)
All questions	523/2870	(18%)	379/2870	(13%)	855/900	(95%)	44/900	(5%)	1/900	(0.1%)

A participant's answer was deemed *statistically guessable* if it was among the five most common answers chosen by all other participants (excluding the participant's partner).