Usable Security
Computer Security Lecture 17

David Aspinall[1]

School of Informatics
University of Edinburgh

15th March 2013

[1]based on slides by Mike Just. Material for two lectures, will be delivered in one!

Outline

Outline

Why Psychology?

Psychology gives us tools to study human bevaviour.

Why Psychology?

Psychology gives us tools to study human bevaviour.

This helps us understand and predict **the user**.

- ▸ *"How do users behave, and why?"*
- ▸ Apply **Human-Computer Interaction (HCI)**, which began as applied psychology.

Why Psychology?

Psychology gives us tools to study human bevaviour.

This helps us understand and predict **the user**.

- ▶ *"How do users behave, and why?"*
- ▶ Apply **Human-Computer Interaction (HCI)**, which began as applied psychology.

It also helps us understand and predict **the attacker**.

- ▶ *"How do attackers behave, and why?"*
- ▶ Emerging area of study; related to **Criminology**.
- ▶ Study of behaviour of hackers, countermeasures and responses (including by society)

In this lecture we concentrate mainly on **the user**.

Human versus Computer

Security and Usability: First thoughts

- Usable security ≡ Nice GUI?

- Secure **and** usable technology?
  - Are you joking?
  - Surely secure means hard-to-use!

- Or: software security is usable. What's the problem?
  - Maybe for a technical person . . .

Security and Usability: Odd Finding

- Nice GUIs are necessary, but not sufficient for usable security.

- Though, sadly, many users believe the converse: that a nice GUI implies security!

> *The web site that was judged to have the best presentation as determined by [. . .] participants' ratings was the site judged to be the most secure.*
>
> *[Carl Turner, How do consumers form their judgement of the security of e-commerce web sites?, April 2003.]*

Security and Usability: A Tradeoff?

Does increased security decrease usability?

- A 20-character password doesn't seem very usable

Security and Usability: A Tradeoff?

Does increased security decrease usability?

- A 20-character password doesn't seem very usable
- **But** an easier way to recall information of equivalent strength (e.g., a graphical password) might be more usable

Security and Usability: A Tradeoff?

Does increased security decrease usability?

- A 20-character password doesn't seem very usable
- **But** an easier way to recall information of equivalent strength (e.g., a graphical password) might be more usable

Security and Usability: A Tradeoff?

Does increased security decrease usability?

- ► A 20-character password doesn't seem very usable
- ► **But** an easier way to recall information of equivalent strength (e.g., a graphical password) might be more usable

Does increased usability decrease security?

- ► Clearly, a 1-character password is easy to recall, though not secure.

Security and Usability: A Tradeoff?

Does increased security decrease usability?

- ▶ A 20-character password doesn't seem very usable
- ▶ **But** an easier way to recall information of equivalent strength (e.g., a graphical password) might be more usable

Does increased usability decrease security?

- ▶ Clearly, a 1-character password is easy to recall, though not secure.
- ▶ **But** allowing you to store your 8-character password on a token might be more usable than remembering it.

Security and Usability: A Tradeoff?

Does increased security decrease usability?

- A 20-character password doesn't seem very usable
- **But** an easier way to recall information of equivalent strength (e.g., a graphical password) might be more usable

Does increased usability decrease security?

- Clearly, a 1-character password is easy to recall, though not secure.
- **But** allowing you to store your 8-character password on a token might be more usable than remembering it.

Security and Usability: A Tradeoff?

Does increased security decrease usability?

- ▶ A 20-character password doesn't seem very usable
- ▶ **But** an easier way to recall information of equivalent strength (e.g., a graphical password) might be more usable

Does increased usability decrease security?

- ▶ Clearly, a 1-character password is easy to recall, though not secure.
- ▶ **But** allowing you to store your 8-character password on a token might be more usable than remembering it.

Q. What about decreased usability?

State of the Nation

- Many security systems are not usable; self-defeating
  - Decrease productivity (obstacles to task completion)
  - Promote insecure behaviour (e.g., writing down passwords)

- Attackers exploit usability flaws
  - Bad, inconsistent interfaces $\longrightarrow$ opportunities

- Security technology is often avoided
  - Too complicated to use or implement
  - Benefits not well understood
  - Viewed as an obstacle to productivity
  - Really not needed, or no RoI/risk assessment

Outline

Human-Computer Interaction (HCI)

Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them. [ACM]

Evolution and components of HCI:

- Earlier fields of Human Factors & Ergonomics
- User Interface Design
- User/Human Centred Design
  - Engineering with user focus at all stages
  - Participatory design explicitly includes users
- Interaction Design, User-Experience (UX)
  - More emphasis on cognitive/experiential factors

The Environment
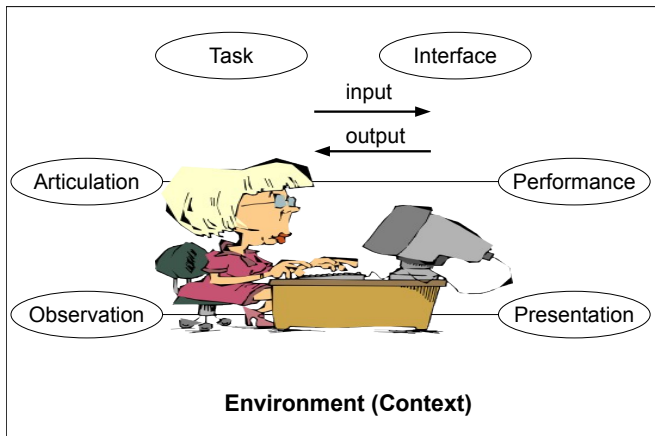
The Environment – User Characteristics

Physical

- ► Characteristics and limitations of the human body
- ► Quality of characteristics varies, e.g., biometrics
- ► Ageing and illness make some tasks difficult, or at least time consuming
- ► Accessibility can be a major consideration

The Environment – User Characteristics

Physical

- ▶ Characteristics and limitations of the human body
- ▶ Quality of characteristics varies, e.g., biometrics
- ▶ Ageing and illness make some tasks difficult, or at least time consuming
- ▶ Accessibility can be a major consideration

Mental

- ▶ Memory: learned behaviour
- ▶ Perceptions: mental models, anticipated behaviour
- ▶ Attitudes and beliefs: e.g., valid metaphors, diligence levels

The Environment – Tasks

Humans are **goal-oriented**

- ▶ behaviour when completing tasks shows this
- ▶ e.g., early ATM designs: card-after-money ⟶ many lost cards left at machines

But security is often **external** to main goal

- ▶ Smooth integration becomes a key part of design
- ▶ So: want to minimise time, number of interactions
- ▶ But likely to *intervene* or *precede* rather than follow.

The Environment – Context

Physical
- Atmosphere
- Climate
- Lighting, noise
- Desk, chair, desktop
- Mobile versus fixed

Social
- Private, semi-private, public
- Social norms regarding acceptable user behaviour
- Organizational culture, e.g., professional behaviour

Classical system design with HCI

Iteration within the waterfall model:
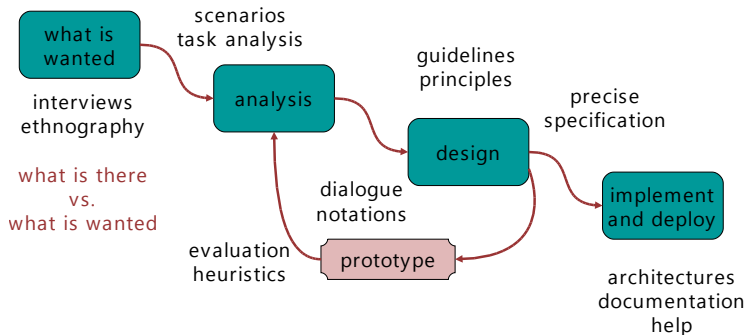
[Dix et al, *Human Computer Interaction*, 3rd Ed, 2004, p.195.]

HCI Design for Security

For security, a number of principles stand out:

- **Interface**
  - User control and freedom
  - Consistency and standards
  - Flexible and efficient
  - Informative feedback
- **Errors**
  - Design for error (e.g., safe defaults, non-tech messages)
  - Help with error recognition and recovery (reversal)
- **Memory**
  - Memory recognition over recall
  - Memory *in the world* (not just *in the head*)

Dimensions for Evaluating Designs

- **Who** is giving the feedback? Design expert? Fellow programmer? A typical user or member of a target user group? One person or a significant proportion?

Dimensions for Evaluating Designs

- **Who** is giving the feedback? Design expert? Fellow programmer? A typical user or member of a target user group? One person or a significant proportion?
- **When** are you getting this feedback? On an early prototype or an established product?

Dimensions for Evaluating Designs

- **Who** is giving the feedback? Design expert? Fellow programmer? A typical user or member of a target user group? One person or a significant proportion?
- **When** are you getting this feedback? On an early prototype or an established product?
- **How** has this evaluation been arrived at? Is it by comparison to some guidelines, or from a simulated walkthrough? Is it from use in a realistic contect ('ecological validity')?

Dimensions for Evaluating Designs

- **Who** is giving the feedback? Design expert? Fellow programmer? A typical user or member of a target user group? One person or a significant proportion?
- **When** are you getting this feedback? On an early prototype or an established product?
- **How** has this evaluation been arrived at? Is it by comparison to some guidelines, or from a simulated walkthrough? Is it from use in a realistic contect ('ecological validity')?
- **What** has been used as a measure? Quantitiative (e.g., time to complete task, error rate) or qualitative (e.g., ease of use ratings)? Compared to recommendations or alternatives? Consistency?

Perspective on Training and Design

systems security is one of the last areas in IT in which user-centred design and user training are not regarded as essential

. . .

hackers pay more attention to the human link in the security chain than security designers do.

[Adams and Sasse, 1999]

This is still true today, but hopefully changing.

Perspective on Expectations

Humans are incapable of securely storing high-quality cryptographic keys, and they have unacceptable speed and accuracy when performing cryptographic operations. (They are also large, expensive to maintain, difficult to manage, and they pollute the environment. It is astonishing that these devices continue to be manufactured and deployed. But they are sufficiently pervasive that we must design our protocols around their limitations.)

[C. Kaufman, R. Perlman, M. Speciner, Network Security, 2002]

Outline

PGP Encrytion

Seminal paper *Why Johnny can't encrypt* by Whitten &
Tygar in 1999.
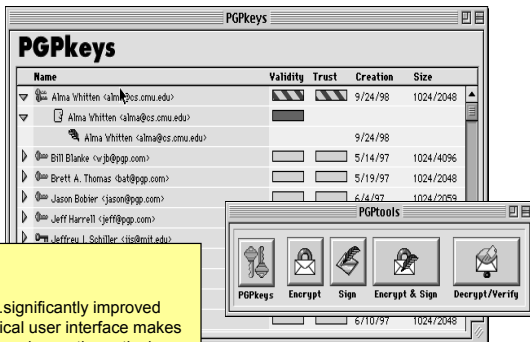
PGP Encrytion (2)

Result: Users had great difficulty.
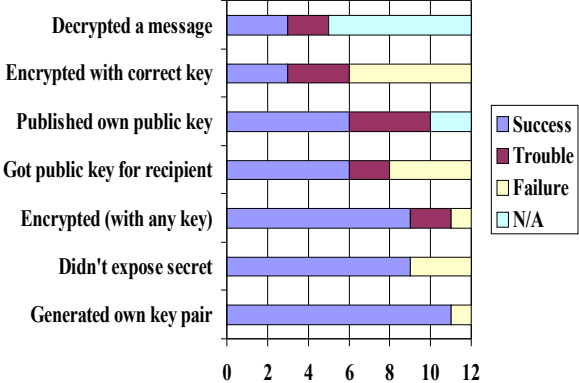
Firewalls (1)

2009 study *Revealing Hidden Context: Improving Mental Models of Personal Firewall Users* by Raja et al, looked at security and usability of Windows Vista firewall.
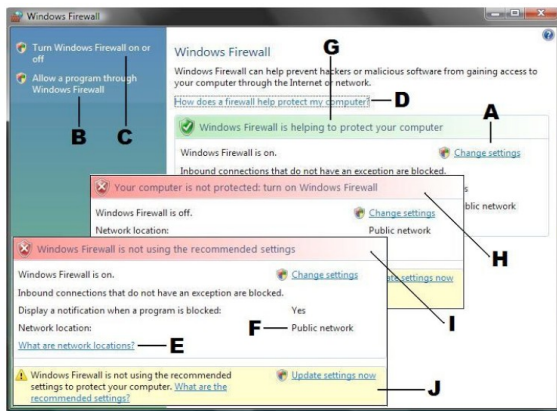
Firewalls (2)

Prototype: reveals hidden context, iteratively designed with pilot users.

Firewalls (3)

Impact of prototype on users' mental models
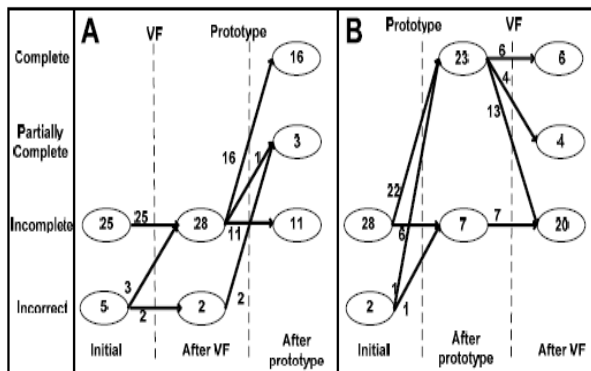
Facemail

- **Facemail** (2007) aimed to mitigate common mistakes:
    - Sending email to a wrong address or list alias
    - Accidentally hitting *Reply-to-all*
    - Resolving multiple users with same name
- These mistakes often cause unintended leaks.
- Proposed mechanism: use images of recipients
- Issues include
    - Collection of images
    - Dealing with mailing lists
    - Dealing with large `To:` or `Cc:` lists

Example – Facemail Interfaces
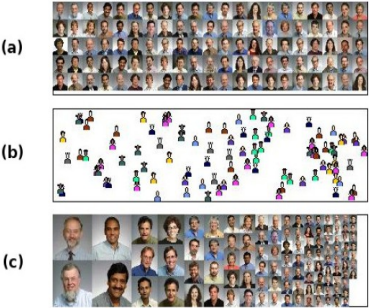
Figure 2: Alternative displays for mailing lists.
(a) Fixed-size grid; (b) Random crowd of icons;
(c) Progressive scaling grid.
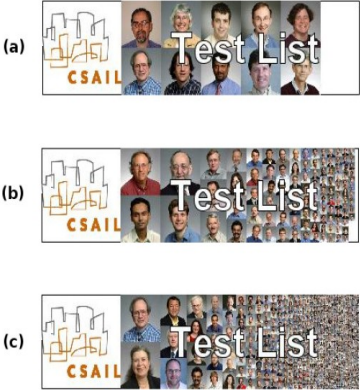
Figure 3: Progressive scaling grid showing (a) 10 faces,
(b) 100 faces, and (c) 1000 faces.

Facemail Results
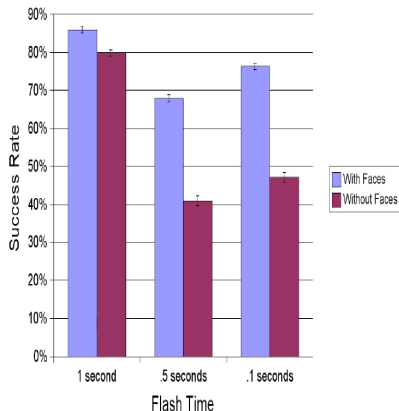
Figure 8: Success rate for the question "Would this email go only to the desired recipient(s)?" Not Sure responses are treated as wrong answers. Error bars show standard error.
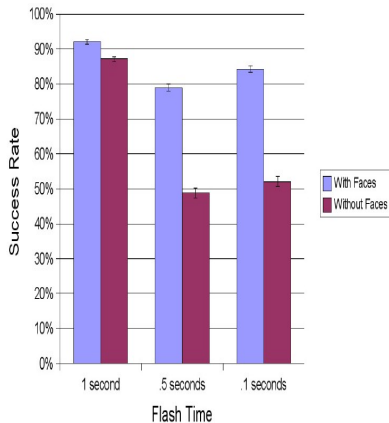
Figure 9: Success rate for the question "How many people would receive this email?" Not Sure responses are treated as wrong answers. Error bars show standard error.

Outline

Usable Authentication

User authentication is the most common and best studied human security task.

- **Information** is used for authentication:
    - something you have, you know, or you are.
    - obviously: varying usability aspects
    - additionally (maybe implicit): location, time

- There is **lifecycle** for auth information:
    - **Issuance**: when information is created or issued
    - **Use**: when information is used to authenticate
    - **Maintenance**: when information is revised/retired

We can consider the usability of each class of information, at each of the stages.

Lifecycle: Something You Have

Issuance
- Requires physical interaction, e.g., mail, in-person
- May require distribution of a reader as well

Use
- Requirements on human memory, e.g., *"Where did I leave my card?"*, *"Which card do I use?"*
- Human-card-machine interface issues, e.g., *"Which way do I insert the card?"*

Maintenance
- Subject to wear-and-tear, loss, theft, forgery
- Require periodic replacements

Lifecycle: Something You Are

Issuance
- ▸ "Reverse issuance" required to submit biometrics
- ▸ Accessibility: not all humans have readable fingerprints, irises, etc.

Use
- ▸ Minimal requirements placed upon human memory, e.g., *"Which finger did I use?"*, though may be specified at authentication
- ▸ Human-machine interface issues, e.g., cut finger

Maintenance
- ▸ Limited options for renewal due to finite set of biometrics

Lifecycle: Something You Know

Issuance

- ▶ Memorizing something new, or selecting already-known
- ▶ Ability to follow rules: length, capitalization, . . .
- ▶ Needs to be sufficiently secure and memorable

Use

- ▶ Can I recall my password (with 100% accuracy)?
- ▶ If so, which one (out of many I have)?
- ▶ If so, which one (out of many updates I've made)?

Maintenance

- ▶ Subject to loss/expiration
- ▶ Re-issuance may require secondary mechanism
- ▶ Rules on freshness/variation for re-issued data

Lifecycle: General Guidelines

Issuance:
- Limit amount of physical interaction
- Limit human processing and learning requirements
- Limit number of seemingly artificial constraints

Use:
- Limit memory requirements
- Limit requirements for perfect accuracy

Maintenance
- See 'Issuance'
- Limit excessive update requirements

Password Authentication

Encryption of password p → Auth DB

password p

App Server ↔ Auth Server

Server
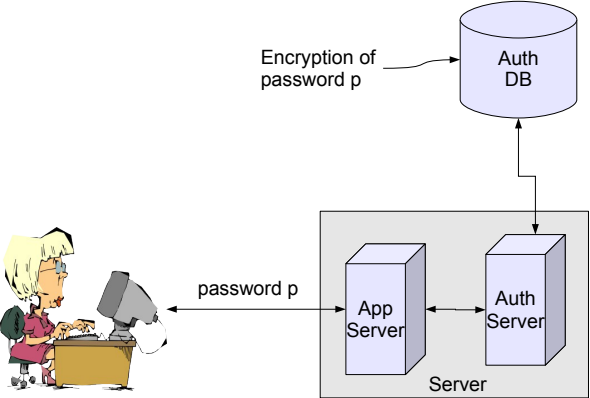
Passwords Lifecycle
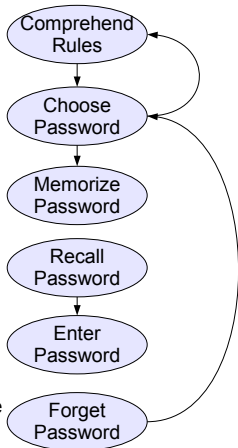
Password Usability Issues

Passwords require "100% correct, unaided recall of a non-meaningful item" [Sasse, 2003]

The cause of usability difficulties are well-known:

- Rules, rules, rules!
  - Length, e.g., of at least 8 characters
  - Diversity, e.g., one uppercase, one lowercase, one number, . . .
  - Update requirements, e.g., monthly
- Limited number of login attempts, e.g., 3
- Dozens of accounts with passwords

Password Recall: Challenges

- Limited capacity of working memory
- Items stored in memory decay over time
- Frequent recall improves memorability of items (automatically)
- Unaided recall is harder than cued recall
- Non-meaningful items are harder to recall than meaningful ones
- Similar items compete and are easily confused
- Items linger in memory – humans cannot "forget on demand"

Password Recall: Some Results

- Causes of password login failure:
  - 52% "memory failure" (old password 37%, wrong system 15%)
  - 20% of users forgot their user identifier
  - 12% mis-typed (typos or hitting Enter too soon)
- Password selection
  - 28% of users' passwords are identical
  - 68% use one method to contruct their passwords
  - 51% of are a word with an appended number
- Password maintenance
  - 90% of users will only change when forced
  - 45% change only by a number increment
- Password "storage"
  - 30% of users write down all their passwords
  - 32% of users write down infrequently used passwords

[Sasse et al, 2000, 2001]

Passwords: Other Security Issues

- Using the same password across multiple accounts
  - Not always a problem, e.g., same password for news subscriptions (marketing)
  - But problematic to mix across high and low risk environments, e.g., work account and news subscription

Passwords: Other Security Issues

- Using the same password across multiple accounts
  - Not always a problem, e.g., same password for news subscriptions (marketing)
  - But problematic to mix across high and low risk environments, e.g., work account and news subscription

- Sharing of passwords or PINs
  - Where assistance is required, e.g., elderly people
  - Where trust may be misplaced, e.g., at the pub

Passwords: Tradeoffs

- Password selection
  - Longer improves security, but strains memory
  - Rules increase search space, but strain human capabilities (and we will take the path of least resistance)
  - Left to our own devices, humans will typically choose weak passwords
- Password use
  - Humans recognize reasonably well, but full recall is sometimes challenging (especially for a password that is external to our main tasks)
  - Humans are overly trusting, and poor judges of risk
- Password maintenance
  - We have trouble forgetting (yet recall isn't precise)
  - With forced update, we take the path of least resistance

Potential Improvements

- More flexible update policies
  - Warn users of impending updates
  - Alternate solutions – "Last login time: "
- Feedback
  - Tell users why their choices are bad
  - Now common: "entropy meter" and dictionary check
- Flexible password storage
  - Is it really that bad to write down your password?
  - Bruce Schneier [2000]: *"Security departments have a knee-jerk reaction against writing passwords down."*
- Increased tolerance
  - More than 3 attempts (Sasse et al. suggest 10, and showed a 50% improvement in login success)
  - Feedback: inform when CAP L0CKS is on
- Training and education
- Interaction

Password Mnemonics

- A Cambridge study (2004) compared:
  1. **User-chosen passwords**: Passwords chosen by users, based upon traditional guidance
  2. **Password mnemonics**: Deriving a password from a phrase, such as the phrase *An apple a day keeps the doctor away*, to give the password Aaadktda.
  3. **Random passwords**: randomly generated
- Experiments on 4 groups of 100 university students

Password Mnemonics

- A Cambridge study (2004) compared:
  1. **User-chosen passwords**: Passwords chosen by users, based upon traditional guidance
  2. **Password mnemonics**: Deriving a password from a phrase, such as the phrase *An apple a day keeps the doctor away*, to give the password Aaadktda.
  3. **Random passwords**: randomly generated
- Experiments on 4 groups of 100 university students
- Unsurprising results:
  - User-chosen: recalled better, but easy to guess
  - Random: recalled poorly, but hard to guess

Password Mnemonics

- A Cambridge study (2004) compared:
  1. **User-chosen passwords**: Passwords chosen by users, based upon traditional guidance
  2. **Password mnemonics**: Deriving a password from a phrase, such as the phrase *An apple a day keeps the doctor away*, to give the password Aaadktda.
  3. **Random passwords**: randomly generated
- Experiments on 4 groups of 100 university students
- Unsurprising results:
  - User-chosen: recalled better, but easy to guess
  - Random: recalled poorly, but hard to guess
- Perhaps suprising results:
  - Mnemonic: can be as hard to guess as random
  - Mnemonic: can be recalled as well as user-chosen
- Mnemonic passwords are sometimes recommended but have so-far failed to catch on widely

Challenge Question Authentication

- **Challenge questions** are a form of authentication credential
- Actually consist of both a *question* and a corresponding *answer*
  - Question and answer are chosen at registration
  - At authentication, the question is presented, the answer solicited
- Often secondary authentication, when password forgotten
  - Recently, used as a complement to passwords
- Use *already known* rather than *specially memorized* info
  - ⟶ hopefully **more memorable**

Types of Challenge Questions

Just's classification:

1. Administratively chosen (Fixed)
   - Questions are presented to the user in a fixed list
   - User generally has the choice to select a number of questions from the list

Types of Challenge Questions

Just's classification:

1. Administratively chosen (Fixed)
   - Questions are presented to the user in a fixed list
   - User generally has the choice to select a number of questions from the list

2. User-chosen (Open)
   - The user has complete freedom in constructing their questions
   - Sample questions and/or guidance is often provided to aid the user

Types of Challenge Questions

Just's classification:

1. Administratively chosen (Fixed)
   - Questions are presented to the user in a fixed list
   - User generally has the choice to select a number of questions from the list

2. User-chosen (Open)
   - The user has complete freedom in constructing their questions
   - Sample questions and/or guidance is often provided to aid the user

3. Hybrid
   - Combination of administratively and user chosen
   - E.g., A partial question with "blanks" filled by user
   - E.g., A general question for which user can contribute additional specifics

A Usability and Security Study

In 2009, Just and Aspinall conducted a study with University of Edinburgh students. We designed an ethical evaluation mechanism:
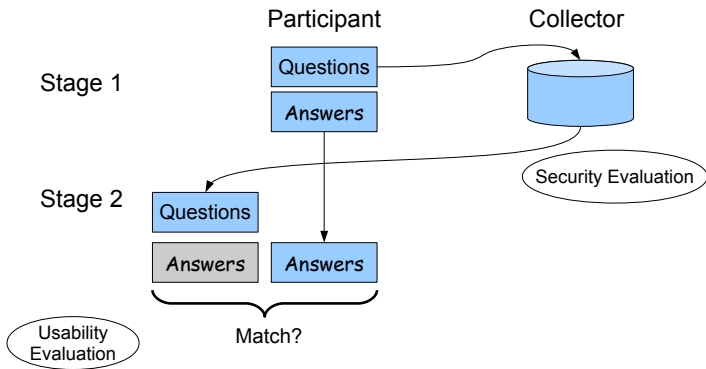
Challenge Question Usability Assessment

- ▶ Criteria based on question structure:
    1. **Applicability**: What proportion of users have sufficient information to choose an answer? Not everyone can answer _"What was my first pet's name?"_
    2. **Memorability**: How well can users recall the original answer (over time)?
    3. **Repeatability**: Can users precisely repeat the original answer? Problems may be _syntactic_ such as spelling, or _semantic_, such as favourites, which may change over time.
- ▶ Measurement:
    - ▶ **Subjectively**. E.g., expected spelling challenges, such as with _"What was my first address?"_ and variantssuch as _Street_, _St_, etc.
    - ▶ **Statistically**. E.g., based on number of pet owners.
    - ▶ **Empirically**. Via experiments, for a particular _user base_: how easy it was for users to choose and recall questions.

Challenge Questions: Usability Results

- Despite using personal knowledge *already known* to users, experiments have shown relatively poor memorability and repeatability results

- Results of 10% - 25% of failed authentication
  - Both for administratively and user chosen questions
  - Even for young participants, with presumably better memories

- Possible reasons:
  - As with passwords, difficulty with precise recall
  - In some cases, users register "false" answers

Challenge Questions: Security Results

- ► Criteria for measurement:
  1. **Blind Guess**: The attacker has no initial information. Their attack success is related to the *answer length*.
     Results: low security, average 7.5 characters.

Challenge Questions: Security Results

- Criteria for measurement:
  1. **Blind Guess**: The attacker has no initial information. Their attack success is related to the *answer length*.
     Results: low security, average 7.5 characters.

  2. **Focused (Statistical) Guess**: The attacker additionally knows the challenge questions. Their attack success is related to the *size of the answer space*.
     Results: low-medium security: many uneven/small distributions

Challenge Questions: Security Results

- Criteria for measurement:
  1. **Blind Guess**: The attacker has no initial information. Their attack success is related to the *answer length*.
     Results: low security, average 7.5 characters.

  2. **Focused (Statistical) Guess**: The attacker additionally knows the challenge questions. Their attack success is related to the *size of the answer space*.
     Results: low-medium security: many uneven/small distributions

  3. **Targeted Observation**: The attacker additionally knows information about the user. Success related to the *availability of the information*.
     Results: low-medium (by self-assessment).

- Recommendation: use multiple questions that are secure along different dimensions

Graphical Authentication

Many alternative mechanisms being researched, sold, even deployed.

For example:

- **Biometrics**: always expected as final answer, yet many difficult practical issues
- **Digital Objects**: use knowledge in-the-world, e.g., basing password on hash of a known image from the web
- **Graphical Authentication**: many schemes
  - Recognition based (e.g., Passfaces, Déjà Vu, Use Your Illusion)
  - Position based (e.g., Passlogix, PassPoints)
  - Action based (e.g., Draw a Secret)

Some pictures follow. Generally: achieve varying levels of security and usability. Still no compelling alternatives to passwords.

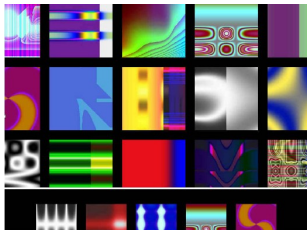# Passlogix Examples
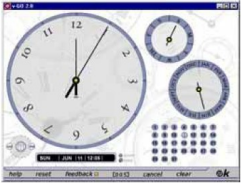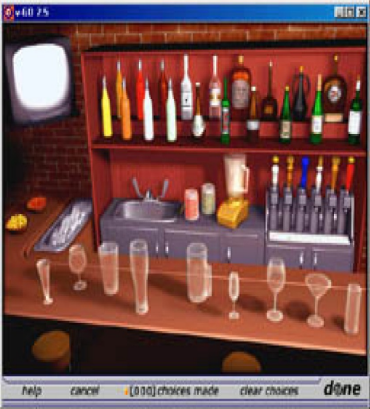
Passspoints Example

Passfaces

A user is required to identify one face from a set of 9 faces. Usually 4 or more stages or panels are used.

Déjà Vu (2)

User picks *n* images from *m* distractors, based on random art images.

Use Your Illusion

Distorted Picture   Original Picture

Authentication and Usability Summary

- A wide variety of solutions for authenticating users
- Possess various security and usability properties
- There is no "silver bullet"
- Some remaining challenges
    - More consistent and comparable evaluations
    - Larger, broader studies
    - Dealing with credential "interference"

Outline

Summary

- Relationship between usability and security
- Introduction to HCI and Usability design
- Discussed many authentication examples to emphasize this relationship:
  - High-level view of different types of authentication information (know, have, are)
  - Password authentication: Traditional and some improvements such as mnemonics
  - Challenge question authentication
  - Many variants of graphical authentication
- Main additional mechanism: **biometrics**

References

- ▶ Since 2005, the *Symposium on Usable Privacy and Security* (SOUPS) has been the main conference for work on both security and usability
  - ▶ http://cups.cs.cmu.edu/soups/

- ▶ Related courses (some material borrowed from these):
  - ▶ Barbara Webb, Human Computer Interaction, University of Edinburgh
  - ▶ Angela Sasse, People and Security, University College London