MEASURING USER ENGAGEMENT

MOUNIA LALMAS, YAHOO! LABS
HEATHER O'BRIEN, UNIVERSITY OF BRITISH COLUMBIA
ELAD YOM-TOV, MICROSOFT RESEARCH

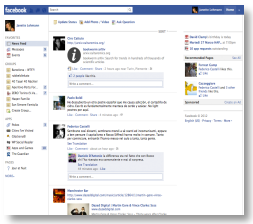
WHY IS IT IMPORTANT TO ENGAGE USERS?

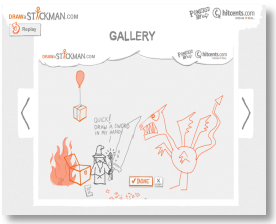
- In today's wired world, users have enhanced expectations about their interactions with technology

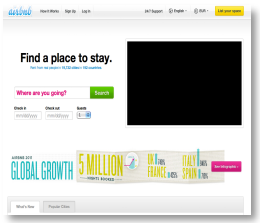
  … resulting in increased competition amongst the purveyors and designers of interactive systems.

- In addition to utilitarian factors, such as usability, we must consider the hedonic and experiential factors of interacting with technology, such as fun, fulfillment, play, and *user engagement*.

- In order to make engaging systems, we need to understand ***what*** user engagement is and ***how*** to measure it.

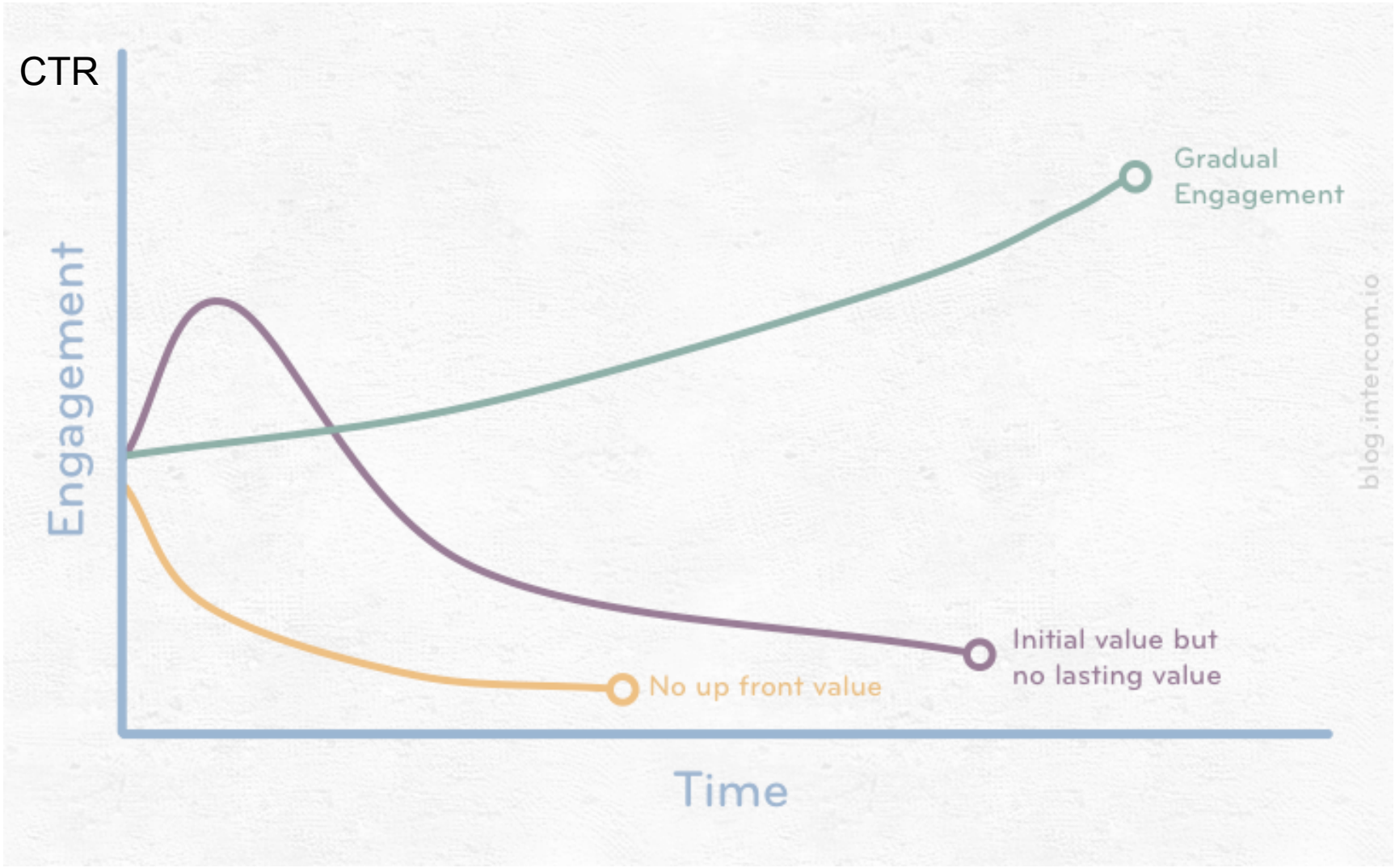






WHY IS IT IMPORTANT TO MEASURE AND INTERPRET USER ENGAGEMENT WELL?

… for example



OUTLINE

- **Introduction and Scope**

- **Part I - Foundations**
  1. Approaches based on self-report measures
  2. Approaches based on web analytics
  3. Approaches based on physiological measures

- **Part II – Advanced Aspects**
  1. Measuring user engagement in mobile information searching
  2. Networked user engagement
  3. Combining different approaches

- **Conclusions**
- **Bibliography**

WHO WE ARE

- **Mounia Lalmas, Visiting Principal Scientist, Yahoo! Labs**
  - Research interest: user engagement, social media, search
  - Blog: http://labtomarket.wordpress.com

- **Heather O'Brien, Assistant Professor, iSchool, University of British Columbia**
  - Research interests: theories of user engagement; self-report and qualitative methods of evaluating user engagement
  - Website: http://faculty.arts.ubc.ca/hobrien/

- **Elad Yom-Tov, Senior Researcher, Microsoft Research**
  - Research interests: learning from user behavior about actions in the physical world
  - Website: http://research.microsoft.com/en-us/people/eladyt/

INTRODUCTION AND SCOPE

ENGAGEMENT IS ON EVERYONE'S MIND

http://iactionable.com/user-engagement/

http://www.localgov.co.uk/index.cfm?method=news.detail&id=109512

http://socialbarrel.com/70-percent-of-brand-engagement-on-pinterest-come-from-users/51032/

.au/article/459294/uses_gamification_drive_user_engagement/

http://www.trefis.com/stock/lnkd/articles/179410/linkedin-makes-a-90-million-bet-on-pulse-to-help-drive-user-engagement/2013-04-15

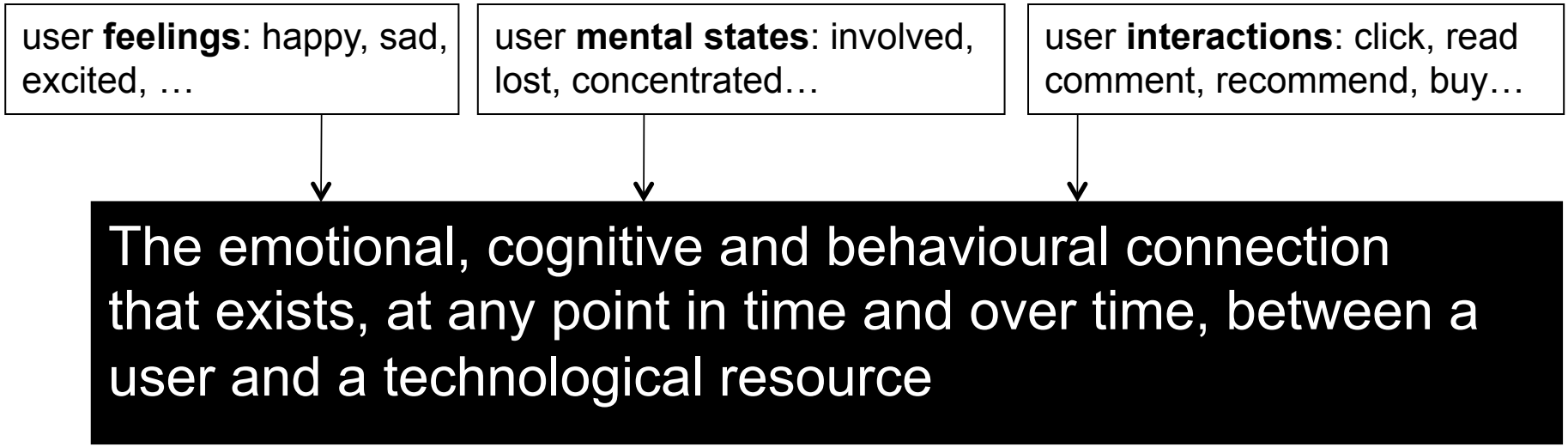
013/05/03/kakao-talk-rolls-out-plus-friend-home-a-ct-users-with-their-favorite-brands/

WHAT IS USER ENGAGEMENT (UE)? (I)

- **"The state of mind that we must attain in order to enjoy a representation of an action" so that we may experience computer worlds "directly, without mediation or distraction"** (Laurel, 1993, pp. 112-113, 116)**.**
- **"Engagement is a user's response to an interaction that gains maintains, and encourages their attention, particularly when they are intrinsically motivated"** (Jacques, 1996, p. 103)**.**
- **A quality of user experience that depends on the aesthetic appeal, novelty, and usability of the system, the ability of the user to attend to and become involved in the experience, and the user's overall evaluation of the experience.** ***Engagement depends on the depth of participation the user is able to achieve with respect to each experiential attribute*** (O'Brien & Toms, 2008)**.**
- **"…explain[s] how and why applications attract people to use them"** (Sutcliffe, 2010, p. 3)**.**

WHAT IS UE? (II)

- **User engagement is a quality of the user experience that emphasizes the positive aspects of interaction – in particular the fact of being captivated by the technology** (Attfield et al, 2011)**.**



INCREASED EMPHASIS ON MEASURING UE

Marketers exploiting secrets of the living brain

By Kelly Crowe, CBC News Posted: Jan 1, 2013 8:15 AM ET | Last Updated: Jan 1, 2013 9:01 AM ET 118

Neuromarketing 3:10

Listen

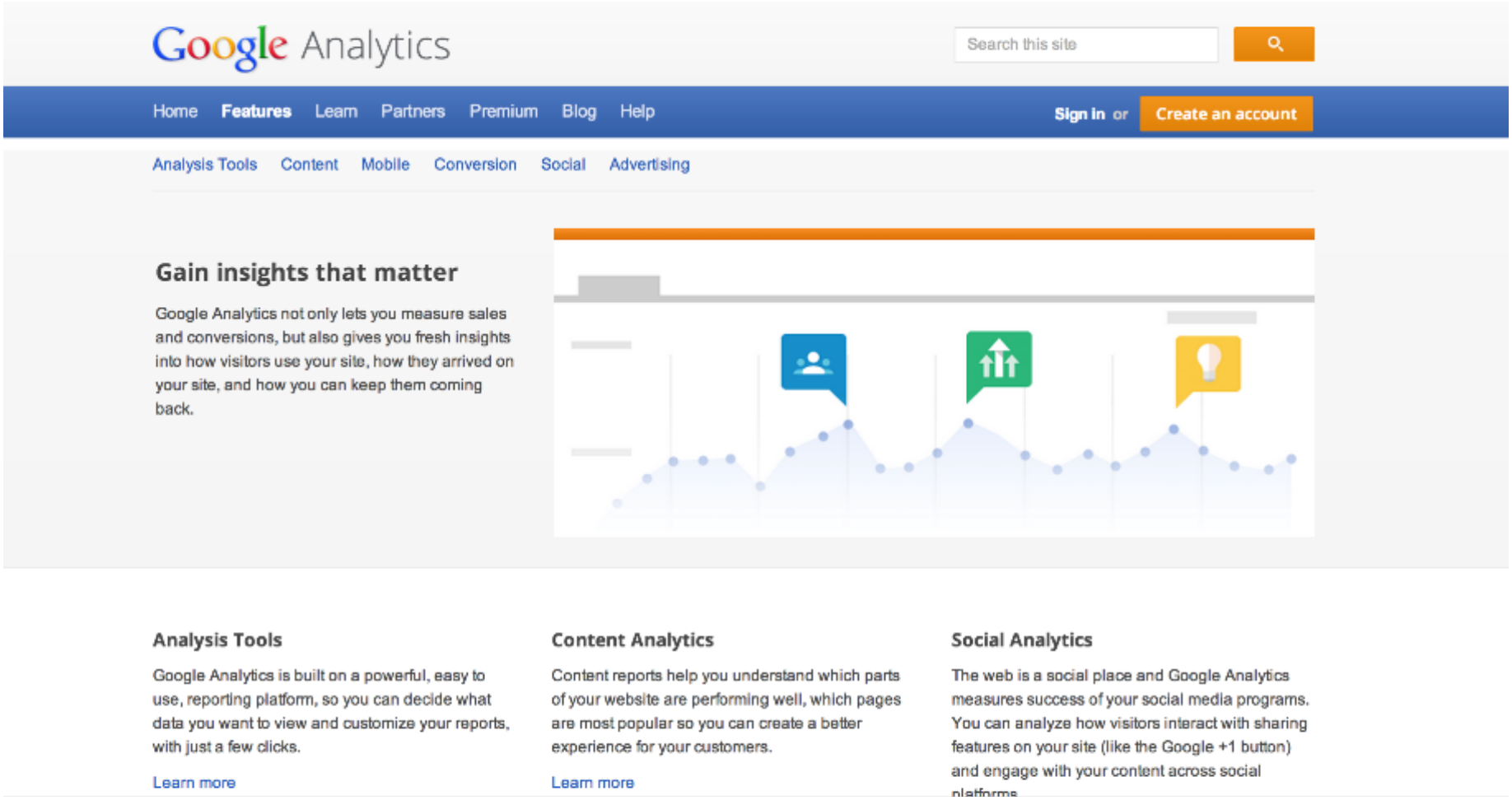
Neuromarketing 1:39

The same primitive impulses that helped early man survive against the evolutionary odds are drawing shopper Denam Drew to a pair of tan suede shoes. At least that's the theory behind neuromarketing, an emerging field that uses the tools of neuroscience to understand the secrets of the consumer brain.

Facebook 376
Twitter 189
+1 37
Share 565
Email

http://www.cbc.ca/news/health/story/2012/12/20/inside-your-brain-neuromarketing.html

TRACKING USER BEHAVIOR



http://www.google.ca/analytics/index.html

HOW DO WE CAPTURE USER ENGAGEMENT?

Bloomberg Businessweek
Technology

Like 109k

Global Economics | Companies & Industries | Politics & Policy | Technology | Markets & Finance | Innovation & Design | Lifestyle

GigaOm

Why Measuring User Engagement Is Harder Than You Think

GIGAOM

Posted on GigaOM
By Mathew Ingram on October 12, 2012 | 0 Comments

Related

Data: 85% of Mobile Traffic But 39% of Revenue—What Gives?

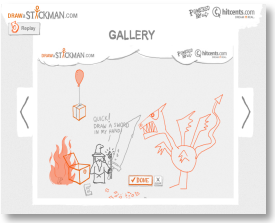
If you've spent any time in a newsroom, traditional or otherwise, you know that publishers are obsessed with measuring where their Web traffic comes from. Whether it's Google (GOOG) Analytics or Chartbeat, or comScore, or Omniture, or any one of a dozen other providers, tracking where readers come from is a crucial part of online media—mostly because publishers need to know which channels are worth focusing on, since there are so many to choose from. Is Twitter your biggest source? Then you should tweet more and optimize your content for Twitter. Is Facebook (FB) a big referrer of

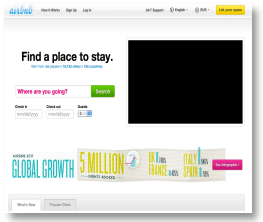
http://www.businessweek.com/articles/2012-10-12/why-measuring-user-engagement-is-harder-than-you-think

WHY IS MEASURING UE IMPORTANT?

- **User engagement is a complex construct**
- **Various approaches have been proposed for measuring engagement, but…**
  - Not enough emphasis on reliability and validity of individual measures, or triangulation of various approaches.
- **Standardization of what user engagement is and how to measure it will benefit research, design, and users.**





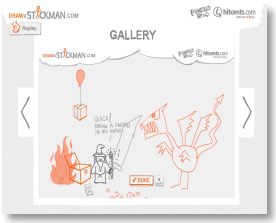


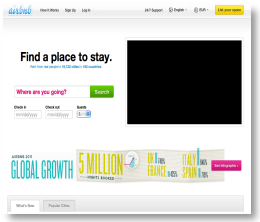
CONSIDERATIONS IN THE MEASUREMENT OF USER ENGAGEMENT

- Short term (within session) and long term (across multiple sessions)
- Laboratory vs. field studies
- Subjective vs. objective measurement
- Large scale (e.g., dwell time of 100,000 people) vs. small scale (gaze patterns of 10 people)
- UE as process vs. as product

One is not better than other; it depends on what is the aim.







SOME CAVEATS (I)

- **This tutorial assumes that web application are "properly designed"**
  - We do not look into how to design good web site (although some user engagement measurement may inform for an enhanced design).
- **This tutorial is based on "published research" literature**
  - We do not know how each individual company and organization measure user engagement (although we guess some common baselines).
- **This tutorial focuses on web applications that users "chose" to engage with**
  - A web tool that has to be used e.g. for work purpose, is totally different (users have no choice).
- **This tutorial is not an "exhaustive" account of all existing works**
  - We focus on work that we came across and that has influenced us; *if we have missed something important, let us know*.

SOME CAVEATS (II)

- **This tutorial focuses on web applications that are widely used by "anybody" on a "large-scale"**
  - User engagement in the game industry or education have different characteristics.
- **This tutorial does not focus on the effect of advertisements on user engagement**
  - We assume that web applications that display ads do so in a "normal" way so that to not annoy or frustrate users.
- **This tutorial looks at user engagement at web application "level"**
  - Although we use examples and may refer to specific sites or types of applications, we do not focus on any particular applications.
- **This tutorial is not about "how" to influence user engagement ☺**

OUTLINE

- **Introduction and Scope**

- **Part I - Foundations**
  1. Approaches based on self-report measures
  2. Approaches based on web analytics
  3. Approaches based on physiological measures

- **Part II – Advanced Aspects**
  1. Measuring user engagement in mobile information searching
  2. Networked user engagement
  3. Combining different approaches

- **Conclusions**
- **Bibliography**

PART 1: FOUNDATIONS

CHARACTERISTICS OF USER ENGAGEMENT (I)

Focused attention
(Webster & Ho, 1997; O'Brien, 2008)

- Users must be focused to be engaged
- Distortions in the subjective perception of time used to measure it

Positive Affect
(O'Brien & Toms, 2008)

- Emotions experienced by user are intrinsically motivating
- Initial affective "hook" can induce a desire for exploration, active discovery or participation

Aesthetics
(Jacques et al, 1995; O'Brien, 2008)

- Sensory, visual appeal of interface stimulates user & promotes focused attention
- Linked to design principles (e.g. symmetry, balance, saliency)

Endurability
(Read, MacFarlane, & Casey, 2002; O'Brien, 2008)

- People remember enjoyable, useful, engaging experiences and want to repeat them
- Reflected in e.g. the propensity of users to recommend an experience/a site/a product

CHARACTERISTICS OF USER ENGAGEMENT (II)

Novelty
(Webster & Ho, 1997; O'Brien, 2008)

- Novelty, surprise, unfamiliarity and the unexpected
- Appeal to users' curiosity; encourages inquisitive behavior and promotes repeated engagement

Richness and control
(Jacques et al, 1995; Webster & Ho, 1997)

- Richness captures the growth potential of an activity
- Control captures the extent to which a person is able to achieve this growth potential

Reputation, trust and expectation (Attfield et al, 2011)

- Trust is a necessary condition for user engagement
- Implicit contract among people and entities which is more than technological

Motivation, interests, incentives, and benefits (Jacques et al., 1995; O'Brien & Toms, 2008)

- Difficulties in setting up "laboratory" style experiments
- Why should users engage?

FORRESTER RESEARCH – THE FOUR I'S

Involvement	• Presence of a user • Measured by e.g. number of visitors, time spent
Interaction	• Action of a user • Measured by e.g. CTR, online transaction, uploaded photos or videos
Intimacy	• Affection or aversion of a user • Measured by e.g. satisfaction rating, sentiment analysis in blogs, comments, surveys, questionnaires
Influence	• Likelihood a user advocates • Measured by e.g. forwarded content, invitation to join

(Forrester Research, June 2008)

FLOW: THE THEORY OF OPTIMAL EXPERIENCE

- **What is "Flow"**

  the state in which people are so involved in an activity that nothing else seems to matter; the experience itself is so enjoyable that people will do it even at great cost, for the sheer sake of doing it (Csikszentmihalyi, 1990, p. 4).

- **Engagement has been called "flow without user control" and "a subset of flow"**

  (Webster & Ahuja, 2004, p. 8)

Attributes of flow

Enjoyment , Focused attention, Absorption, Time perception,

Clear goals and feedback, Control (Cskiszentmihalyi, 1990)

RELEVANCE OF FLOW TO ENGAGEMENT

Flow	Engagement
Feedback from an activity	**Perceived usability** vital for engagement to be sustained
Control during an interaction	
Appropriate levels of **challenge**	
Focused attention	Complete absorption not necessary; getting "sidetracked" may be acceptable and engaging
Intrinsic motivation	May be extrinsic; may be more fruitful to explore motivations as **utilitarian** and **hedonic**
Goal-directed behaviour	Have **fun**, have **an experience**; see where the road takes me
Emphasis on the **individual** and **task** variables	Personal and task relevance important, but characteristics of **system** and **content** precipitate engagement

(O'Brien, 2008)

MEASURING USER ENGAGEMENT

	Measures	Characteristics
Self-reported engagement	Questionnaire, interview, report, product reaction cards, think-aloud	Subjective Short- and long-term Lab and field Small-scale Product outcome
Cognitive engagement	Task-based methods (time spent, follow-on task) Neurological measures (e.g. EEG) Physiological measures (e.g. eye tracking, mouse-tracking)	Objective Short-term Lab and field Small-scale and large-scale Process outcome
Interaction engagement	Web analytics *metrics + models*	Objective Short- and long-term Field Large-scale Process

MEASURES

… a bit more about them

Subjective perception of time (Baldauf, Burgarda & Wittmann, 2009)	• Ask a user to make some estimation of the passage of time during an activity.
Physiological measures	• Involuntary body responses • Gaze behavior, mouse gestures, biometrics (e.g., skin conductance, body temperature, blood volume pulse), facial expression analysis
Follow-on task performance (Jennett et al, 2008)	• How well somebody performs on a task immediately following a period of engaged interaction
Online behaviour	• An estimate of the degree and depth of visitor interaction against a clearly defined set of goals • Based on web analytics (e.g. click-through rate, comments posted)
Search (evaluation)	• Relate system effectiveness and user satisfaction • Designing user models is an important and active research area

OUTLINE

- **Introduction and Scope**

- **Part I - Foundations**
  1. Approaches based on self-report measures
  2. Approaches based on web analytics
  3. Approaches based on physiological measures

- **Part II – Advanced Aspects**
  1. Measuring user engagement in mobile information searching
  2. Networked user engagement
  3. Combining different approaches

- **Conclusions**
- **Bibliography**

PART 1: FOUNDATIONS

APPROACHES BASED ON SELF-REPORT MEASURES

INTRODUCTION TO SELF-REPORT MEASURES

- **What are self-report measures?**
  - A type of method commonly used in social science where individuals express their attitudes, feelings, beliefs or knowledge about a subject or situation.

- **Why consider self-reports?**
  - Emphasize individuals' perceptions and subjective experiences of their engagement with technologies.

- **Self-report methods may be discrete, dimensional, and free response.** (Lopatovska & Arapakis, 2011)

ADVANTAGES OF SELF-REPORT MEASURES

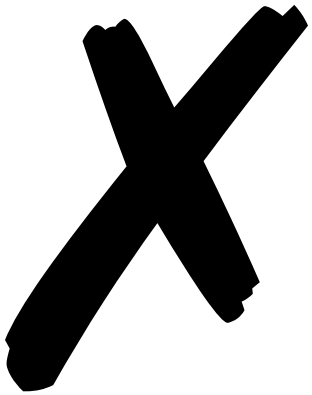
- **Flexibly applied in a variety of settings**
- **High internal consistency for well-constructed measures**
- **Convenient to administer**
- **Specificity in construct definition**
- **Quantitative self-report measures, i.e., questionnaires**
  - Enable statistical analysis and standardization
  - Participant anonymity
  - Administered to individuals or groups
  - Paper-based or web-based
  - Function well in large-sample research studies

(Fulmer & Frijters, 2009)

DISADVANTAGES OF SELF-MEASURES

- **Information processing issues**
  - Interpretation of researchers' questions
  - Developmental challenges associated with age or cognitive ability
- **Communication issues**
  - Wording and response options
  - Rapport between interviewer and interviewee
- **Construct issues**
- **Reliability and validity issues**
- **Participants' responses**
  - What does the "neutral" category mean?
  - *Over-estimate* behavior frequency
  - Reliance on recollection.

(Fulmer & Frijters, 2009; Kobayashi & Boase, 2012)



APPROACHES TO STUDYING USER ENGAGEMENT WITH SELF-REPORT MEASURES – OUTLINE

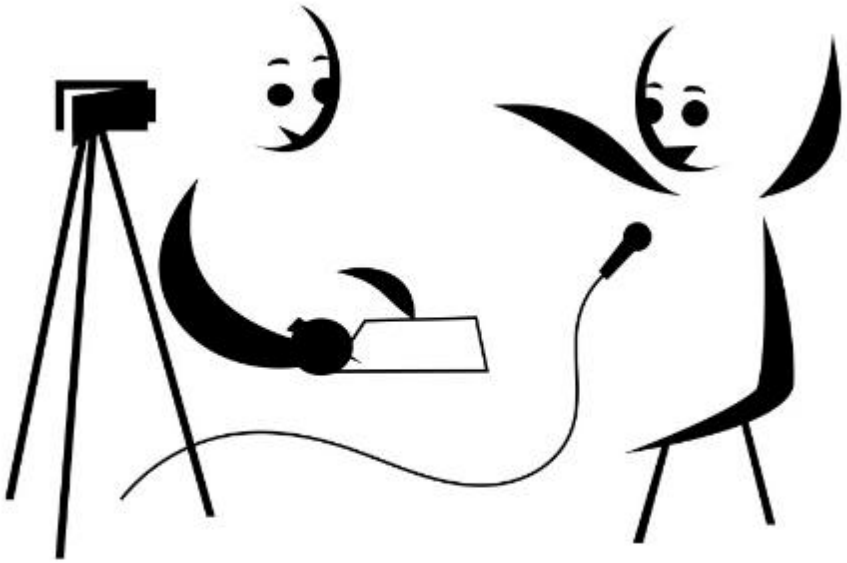
- **Methods**
  - Interviews
  - Think aloud/think after protocols
  - Questionnaires

- **Examples of employing each method to study engagement**

- **Examples of using self-report methods**

INTERVIEWS

- May be structured, semi-structured or unstructured.
- The interview schedule.
- May be one-on-one or one-to-many (focus groups).
- May focus on general or specific events, experiences, or timeframes.



http://openclipart.org/detail/173434/interview-by-jammi-evil-173434

USING INTERVIEWS TO MEASURE USER ENGAGEMENT

- **Objectives:**
  1. To develop an operational definition of engagement, and
  2. To identify key attributes of engagement.

- **Who?**
  - 17 online searchers, gamers, learners and shoppers.

- **Why interviews?**

- **How were the questions formulated?**
  - Grounded in interdisciplinary literature review and theory

- **What guided the analysis?**
  - *Threads of Experience* (McCarthy & Wright, 2004)

(O'Brien & Toms, 2008)

USING INTERVIEWS TO MEASURE USER ENGAGEMENT: OUTCOMES

- **Developed a process-based model of user engagement.**

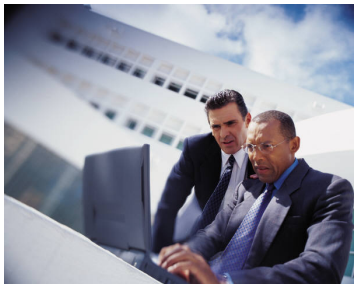
- **Identified attributes of engagement:**
  - Aesthetic and sensory appeal, affect, feedback, control, interactivity, novelty, focused attention, motivation, interest.

- **Mapped attributes to stages in the process model.**

- **Benefit of using interviews.**

(O'Brien & Toms, 2008)

THINK ALOUD/THINK AFTER PROTOCOLS



- **Think aloud**
  - Verbalization *during* the human-computer interaction
- **Think after or simulated recall**
  - Verbalization *after* the human-computer interaction
- **Constructive interaction**
  - Involves two verbalizing their thoughts as they interact with each other
- **Spontaneous and prompted self-report**
  - Participants provide feedback at fixed intervals or at other points defined by the researcher

(Branch, 2000; Ericson & Simon, 1984; Kelly, 2009; Van den Haak, De Jong, & Schellens, 2009)

THINK ALOUD/THINK AFTER PROTOCOLS: CONSIDERATIONS

- **Automatic processes difficult to articulate.**

- **Complex/highly visual interactions may be challenging to remember and/or verbalize.**

- **Think aloud/spontaneous or prompted self-report**
  - Unnatural, interruptive
  - Increased cognitive load

- **Think after or simulated recall:**
  - Relies on memory but attention is less divided
  - Researcher can draw participants' attention to specific features of the interface, activities, etc.

(Branch, 2000; Ericson & Simon, 1984; Kelly, 2009; Van den Haak, De Jong, & Schellens, 2009)

USING THINK ALOUD TO STUDY USER ENGAGEMENT WITH EDUCATIONAL MULTIMEDIA

- **Series of studies with educational multimedia and television advertisements**

- **Think aloud component of the research:**
  - Identified salient aspects of engagement with content and media
    - Content: Perceptions driven by personal interest
    - Media: Focus on media preference, presentation, and affordances of control in navigation

(Jacques, Preece & Carey, 1995)

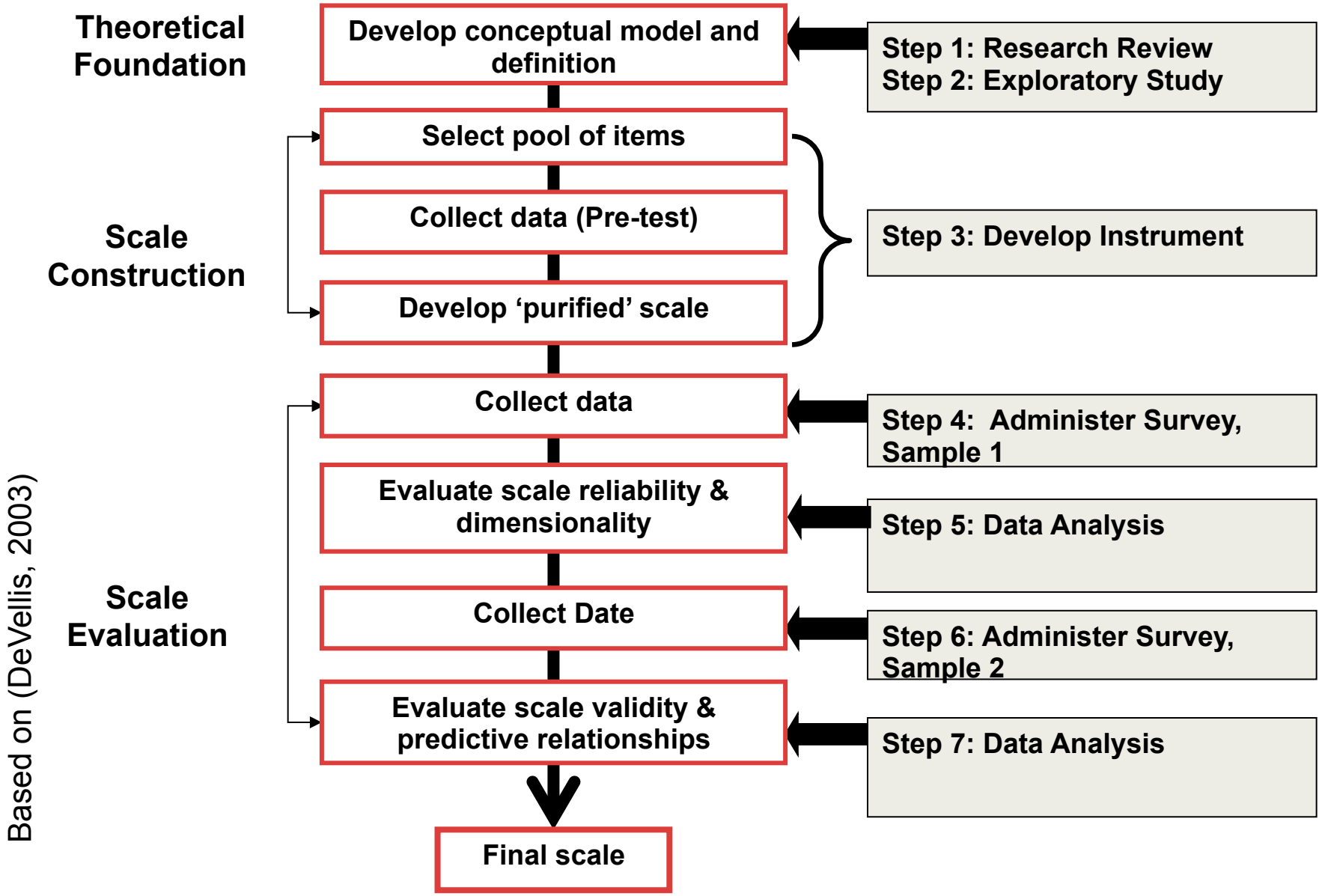
QUESTIONNAIRES

- **Closed-ended (quantitative) and open-ended (qualitative).**

- **Effect of mode** (Kelly et al., 2008)**.**

- **Scale development and evaluation is a longitudinal process.**

SCALE DEVELOPMENT AND EVALUATION



Based on (DeVellis, 2003)

QUESTIONNAIRES FOR MEASURING USER ENGAGEMENT

- ***Jacques, 1996***
  - 13-items
  - Attention, perceived time, motivation, needs, control, attitudes, and overall engagement

- ***Webster & Ho, 1997***
  - 15-items
  - Influences on engagement: including challenge, feedback, control and variety, and
  - Engagement, including attention focus, curiosity, intrinsic interest, and overall engagement.

- ***O'Brien & Toms, 2010*** **– User Engagement Scale (UES)**
  - 31-items
  - Aesthetic appeal, novelty, felt involvement, focused attention, perceived usability, and endurability (overall experience)

USING QUESTIONNAIRES TO STUDY ENGAGEMENT: ROLE OF MEDIA FORMAT: EXAMPLE I

Participants
(*n*=82)

Media Condition
Video
Audio
Narrative text
Transcript text

Pre-task Survey

Story 1

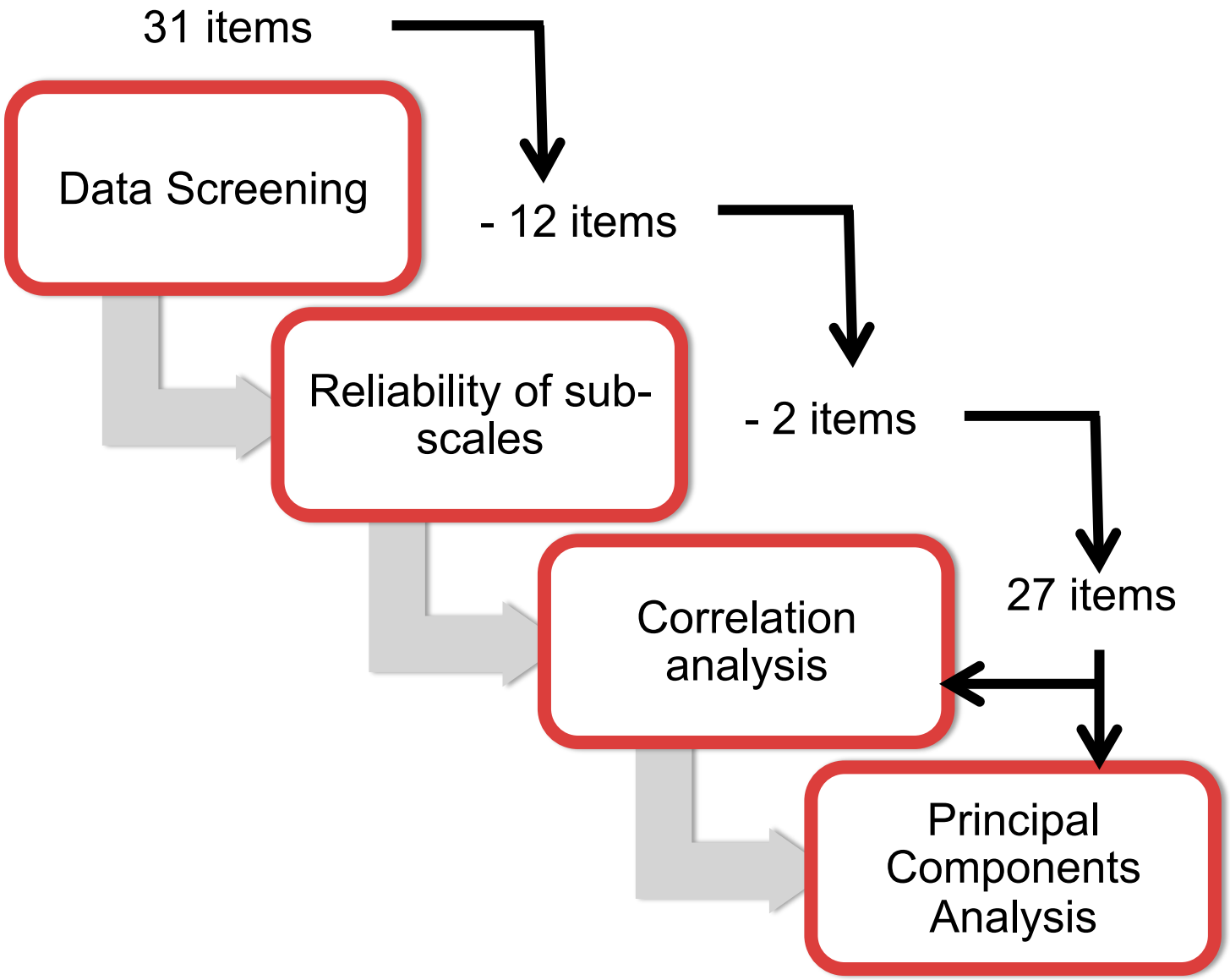
Story 2

UES + Information Recall Questions

Post Session Questionnaire – Attitudes Checklist (Schraw et al. 1998) + Interviews

(O'Brien, 2013)

ROLE OF FORMAT IN MEDIA ENGAGEMENT: PREPARATION AND SCREENING OF UES



PRINCIPLE COMPONENTS ANALYSIS (PCA) OF REMAINING UES ITEMS

Component	Description	No. Items	% Variance	Cronbach's alpha
1	Hedonic Engagement	12	47.9	0.95
2	Focused Attention	4	11	0.87
3	Affective Usability	4	5.9	0.75
4	Cognitive effort	2	4.6	0.83
Kaiser-Meyer-Olkin Measure of Sampling Adequacy = 0.89 Bartlett's Test of Sphericity = x^2=1621.12(231), p<0.001				

FINDINGS FROM THE STUDY

Relationship between Story and Engagement

Component	Story 1: Mining *M(SD)*	Story 2: Farming *M(SD)*
Hedonic Engagement	4.06 (1.3)	5.06 (1.05)
Focused Attention	3.3 (1.4)	3.93 (1.3)
Affective Usability	4.69 (1.3)	5.6 (0.9)
Cognitive Effort	4.19 (1.5)	5.29 (1.3)

Relationship between Media Condition and Engagement

Component	Audio *M(SD)*	Video *M(SD)*	Transcript *M(SD)*	Narrative *M(SD)*
Hedonic Engagement	4.7(1.2)	**5(1.1)**	3.9(1.4)	4.5(1.2)
Focused Attention	3.6(1.4)	**3.8(1.4)**	3.5(1.4)	3.5(1.5)
Affective Usability	5(1.2)	**5.4(1.1)**	4.9(1.3)	5(1.2)
Cognitive Effort	4.5(1.6)	**5.5(1.1)**	4.1(1.5)	4.8(1.4)

FINDINGS FROM THE STUDY (CONTINUED)

Multivariate Tests for Story and Condition

Effect	Λ	*F*	*df*(1)	*df*(2)	*p*
Story	0.8	05.45	1	98	**.001**
Condition	0.78	1.81	3	98	**.04**
Story x Condition	0.92	0.54	3	98	**.88**

Significant F-tests for Univariate Follow-up

UES Component	Effect	MS	*F*	*df*(1)	*df*(2)	*p*
Hedonic Engagement	Story	14.05	9.95	1	98	**.002**
Focused Attention	Story	10.32	4.78	1	98	**.031**
Affective Usability	Story	23.76	17.71	1	98	**.000**
Cognitive Effort	Story	20.02	11.4	1	98	**.000**
Cognitive Effort	Condition	7.23	4.11	3	98	**.009**

CONCLUSIONS: MEDIA FORMAT AND ENGAGEMENT

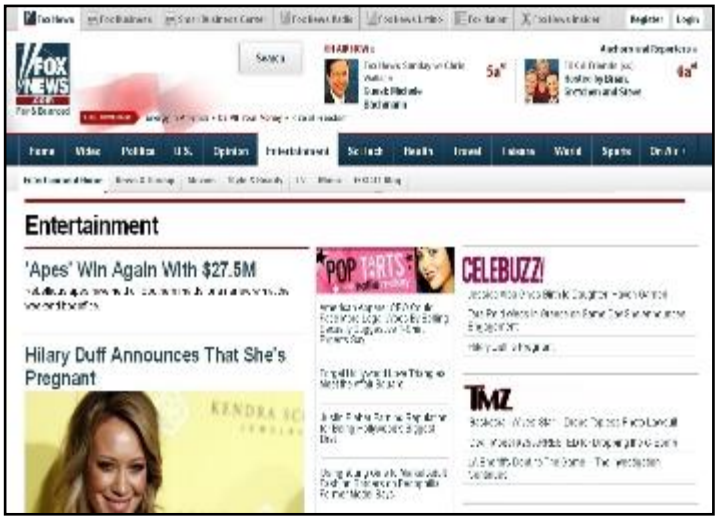
- Next steps in data analysis.
- Value of screening and examining the reliability and principal component structure of the UES items.
- Why performance measures would not be significant in this controlled study.
- What was learned about users' perceived engagement in this study.

EMPLOYING MULTIPLE SELF-REPORT METHODS: EXAMPLE II

- How the visual catchiness (saliency) of "relevant" information impacts user engagement metrics such as focused attention and emotion (affect)
  - **focused attention** refers to the exclusion of other things
  - **affect** relates to the emotions experienced during the interaction
- Saliency model of visual attention developed by (Itti & Koch, 2000)
- Experimenting with crowdsourcing

MANIPULATING SALIENCY

non-salient condition



salient condition

Web page screenshot **Saliency maps**

(McCay-Peet et al, 2012)

STUDY DESIGN

- 8 tasks = finding latest news or headline on celebrity or entertainment topic
- Affect measured pre- and post- task using the Positive e.g. “determined”, “attentive” and Negative e.g. “hostile”, “afraid” Affect Schedule (PANAS)
- Focused attention measured with 7-item *focused attention* subscale e.g. “I was so involved in my news tasks that I lost track of time”, “I blocked things out around me when I was completing the news tasks” and perceived time
- Interest level in topics (pre-task) and questionnaire (post-task) e.g. “I was interested in the content of the web pages”, “I wanted to find out more about the topics that I encountered on the web pages”
- 189 (90+99) participants from Amazon Mechanical Turk

PANAS (10 POSITIVE ITEMS AND 10 NEGATIVE ITEMS)

- **You feel this way right now, that is, at the present moment**

  [1 = very slightly or not at all; 2 = a little; 3 = moderately;

  4 = quite a bit; 5 = extremely]

  [randomize items]

distressed, upset, guilty, scared, hostile, irritable, ashamed, nervous, jittery, afraid
interested, excited, strong, enthusiastic, proud, alert, inspired, determined, attentive, active

(Watson, Clark & Tellegen, 1988)

7-ITEM FOCUSED ATTENTION SUBSCALE (PART OF THE 31-ITEM USER ENGAGEMENT SCALE)

5-POINT SCALE (STRONG DISAGREE TO STRONG AGREE)

1. I lost myself in this news tasks experience
2. I was so involved in my news tasks that I lost track of time
3. I blocked things out around me when I was completing the news tasks
4. When I was performing these news tasks, I lost track of the world around me
5. The time I spent performing these news tasks just slipped away
6. I was absorbed in my news tasks
7. During the news tasks experience I let myself go

(O'Brien & Toms, 2010)

SALIENCY AND POSITIVE AFFECT

- **When headlines are visually non-salient**
  - users are slow at finding them, report more distraction due to web page features, and show a drop in affect
- **When headlines are visually catchy or salient**
  - user find them faster, report that it is easy to focus, and maintain positive affect

---

- **Saliency is helpful in task performance, focusing/avoiding distraction and in maintaining positive affect**

SALIENCY AND FOCUSED ATTENTION

- Adapted focused attention subscale from the online shopping domain to entertainment news domain

- Users reported "easier to focus in the salient condition" BUT no significant improvement in the focused attention subscale or differences in perceived time spent on tasks

---

- **User interest in web page content is a good predictor of focused attention, which in turn is a good predictor of positive affect**

SELF-REPORTING, CROWDSOURCING, SALIENCY AND USER ENGAGEMENT

- **Interaction of saliency, focused attention, and affect, together with user interest, is complex.**
- **Using crowdsourcing worked!**

- **What next?**
  - include **web page content as a quality** of user engagement in focused attention scale
  - more "**realistic**" user (interactive) reading experience
  - other measurements: mouse-tracking, eye-tracking, facial expression analysis, etc.

(McCay-Peet, Lalmas & Navalpakkam, 2012)

CONSIDERATIONS WHEN EMPLOYING SELF-REPORT MEASURES

- **What is the research question?**

- **What is the most suitable self report method?**

- **How might we use self-report in studies of user engagement?**
  - Gather data explicitly about engagement
  - Other self-report measures may predict, validate, or enrich other measures of engagement

- **Why do self-reports get a bad rap?**

PART 1: FOUNDATIONS

APPROACHES BASED ON WEB ANALYTICS

WEB ANALYTICS

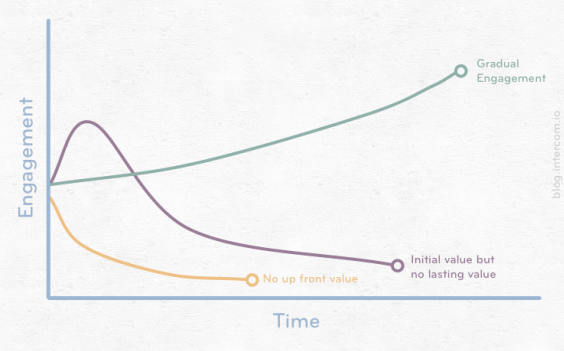
INTRA-SESSION VERSUS INTER-SESSION ENGAGEMENT

- **Intra-session engagement measures our success in attracting the user to remain on our site for as long as possible**
- **"Long-term engagement can be defined as the degree of voluntary use of a system along a wide period of time…"** (Febretti and Garazotto, 2009)
- **Inter-session engagement can be measured directly or, for commercial sites, by observing lifetime customer value (CTR, etc.)**
- **Some studies** (Lehmann et al, 2012) **report some correlation between inter- and intra-session measures, for example, dwell time and number of active days** (ϱ=−0.66)

WHY NOT USE INTRA-SESSION MEASURES EXCLUSIVELY?

- **We seek to have users return to the site again and again, and to perceive the site as beneficial to them**

- **Intra-session measures can easily mislead, especially in for a short time** (Kohavi et al, 2012)**:**
  - Consider a very poor ranking function introduced into a search engine by mistake
  - Therefore, bucket testing may provide erroneous results if only intra-session measures are used



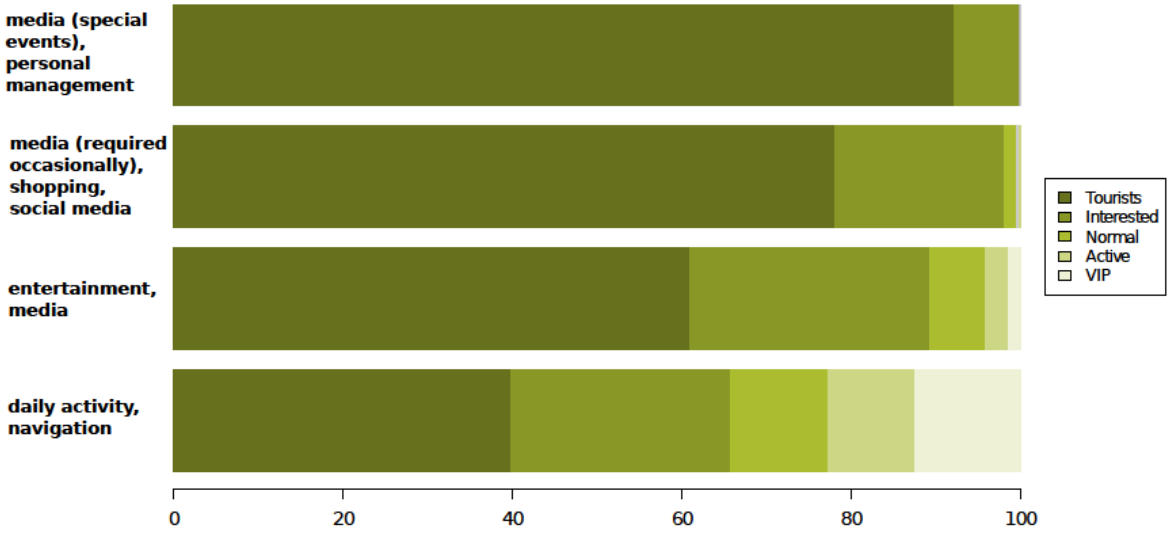
- **Hence inter-session (long-term) engagement is very important**

DEPENDENCY ON USER TYPE

(Lehmann et al, 2012) **observed that different users engage with sites differently.**

Users were defined according to the number of days per month that a site is used:

- **Tourists: 1 day**
- **Interested: 2-4 days**
- **Average: 5-8 days**
- **Active: 9-15 days**
- **VIP: more than 16 days**



Sites were clustered according to the proportion of users from each group.

The figure shows that different sites receive different user types and corresponding usage.

DEPENDENCY ON TASK AND WEBSITE

- **Engagement varies by task. For example, a user who accesses a website to check for emails (a goal-specific task) has different engagement patterns from one who is browsing for leisure.**

- **In one study** (Yom-Tov et al, 2013)**, sessions in which 50% or more of the visited sites belonged to the five most common sites (for each user) were classified as goal-specific.**
  - Goal-specific sessions accounted for 38% of sessions
  - Most users (92%) have both goal-specific and non-goal-specific sessions.
  - The average downstream engagement (more later) in goal-specific sessions was 0.16. This is to be contrasted with 0.2 during non-goal-specific sessions.

- **Dependence on website is clear: news site will see different engagement patterns that online shopping sites.**

LARGE-SCALE MEASUREMENTS OF USER ENGAGEMENT

	Intra-session measures	Inter-session measures
Single site	• Dwell time \ session duration • Play time (video) • Clickthrough rate (CTR) • (Mouse movement) • Number of pages viewed (click depth) • Conversion rate (mostly for e-commerce) • Number of UGC (comments)	• Fraction of return visits • Time between visits (inter-session time, absence time) • Number of views (video) • Total view time per month (video) • Lifetime value (number of actions) • Number of sessions per unit of time • Total usage time per unit of time • Number of friends on site (social networks)
Multiple sites	• Downstream engagement • Revisits (online multitasking)	

ANOTHER CATEGORIZATION OF MEASURES

- (Lehmann et al, 2012) **used a different categorization of measures:**
  - **Popularity**: Total number of users to a site, number of visits, and number of clicks
  - **Activity**: Number of page views per visit, time per visit (dwell time)
  - **Loyalty**: Number of days a user visits a site, number of times visited, total time spent

- **Each of these categories captures a different facet of engagement, and are therefore not highly correlated**

… more about this later

DWELL TIME AND OTHER SIMILAR MEASURES

- **Definition**

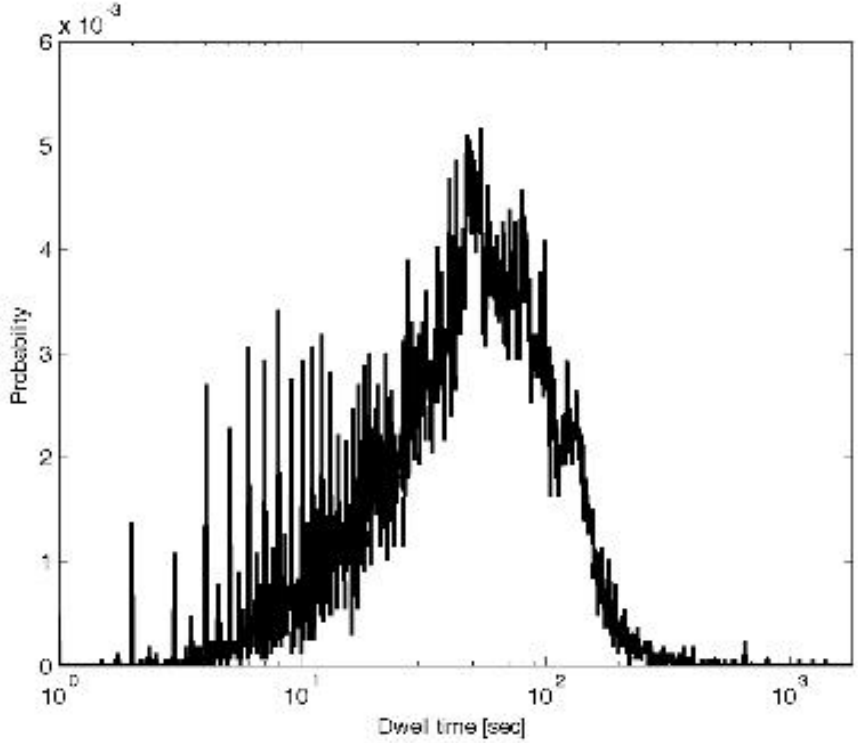
  The contiguous time spent site or web page

- **Similar measures**

  Play time (for video sites)

- **Cons**

  Not clear that the user was actually looking at the site while there



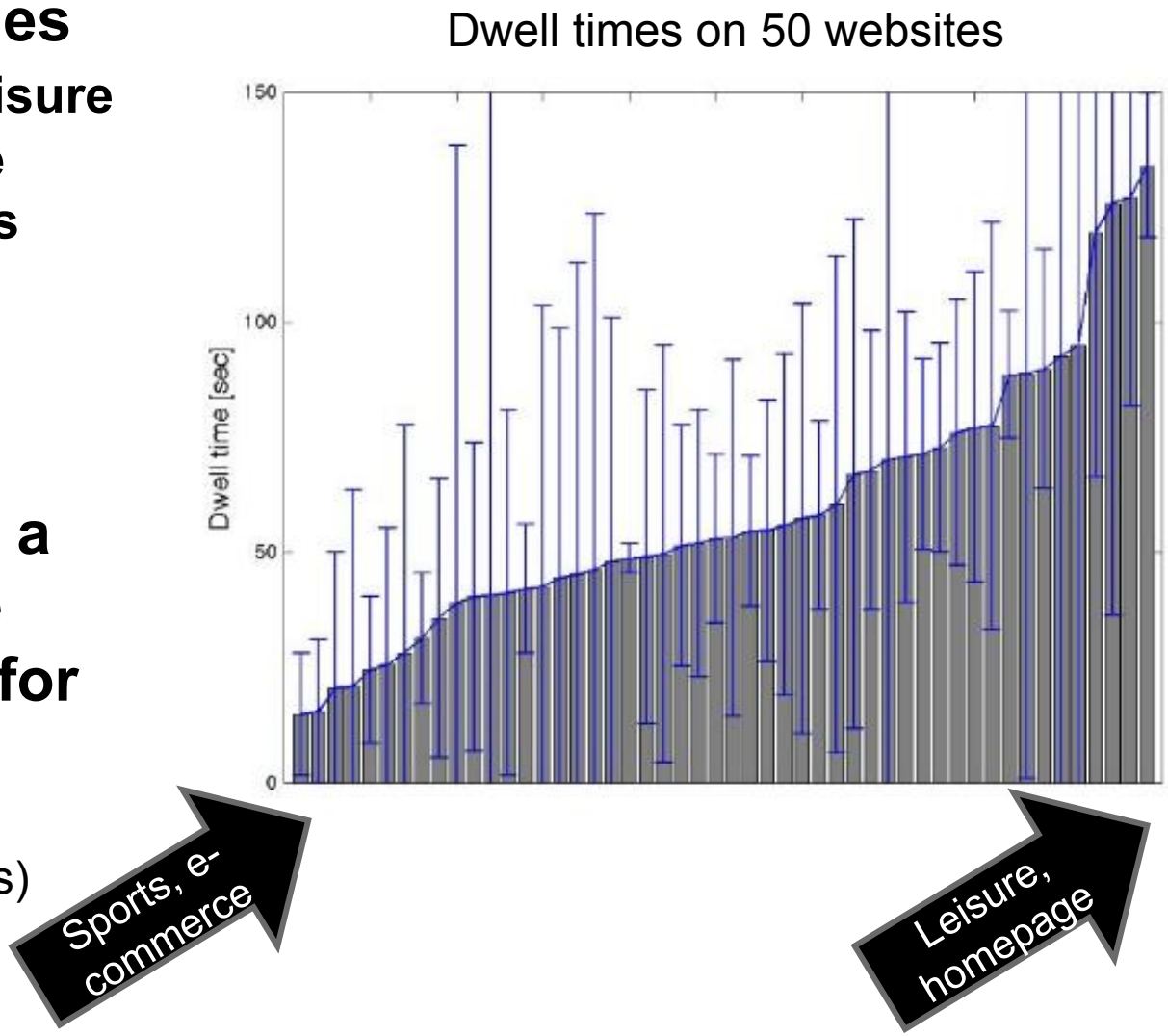
Distribution of dwell times on 50 websites

DWELL TIMES

- **Dwell time varies by site type: leisure sites tend to have longer dwell times than news, ecommerce, etc.**

- **Dwell time has a relatively large variance even for the same site**

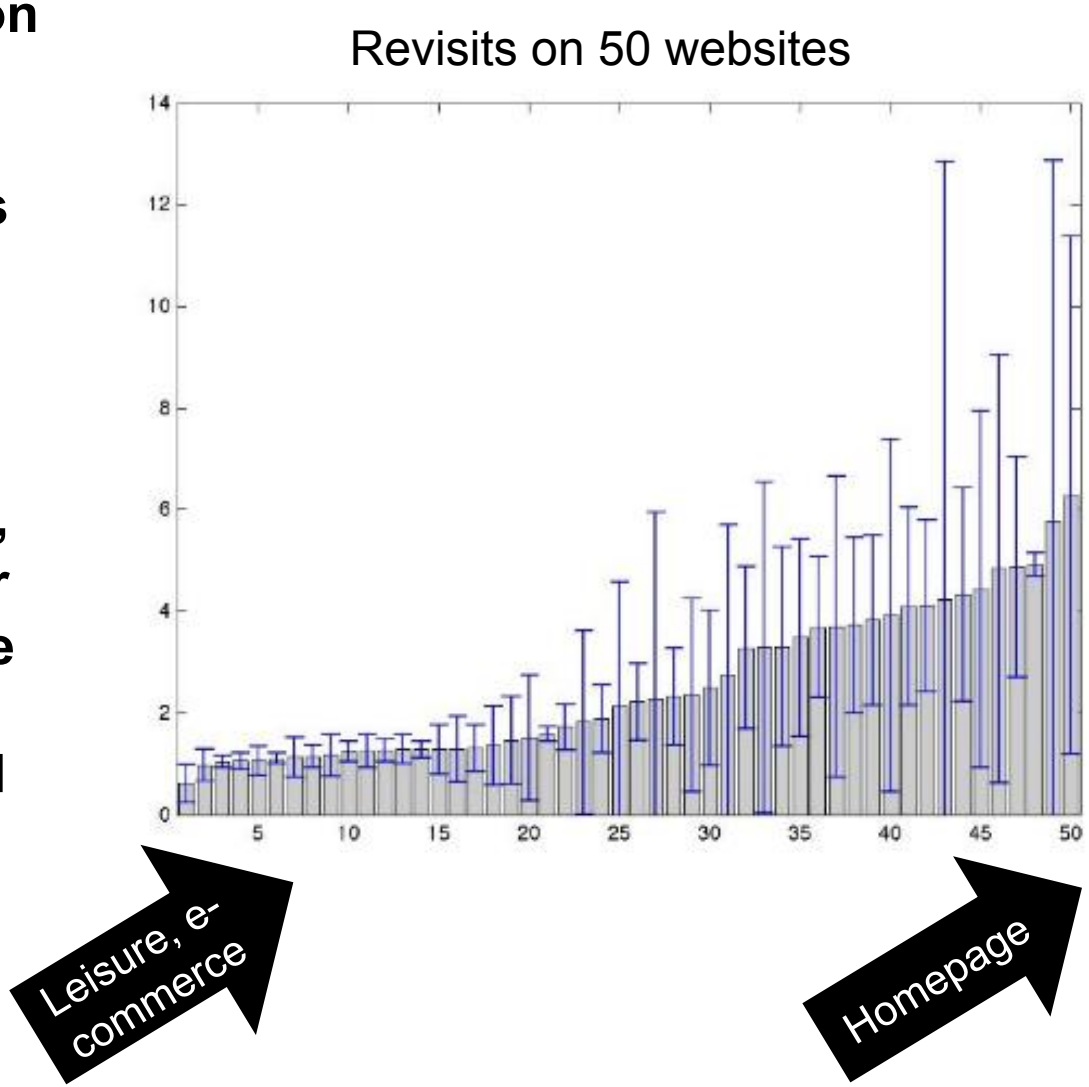
  (recall tourists, VIP, active … users)



REVISITS TO A SITE (WITHIN SESSION)

- **User revisits are common in sites which may be browser homepages, or contain content which is of regular interest to users.**

- **Goal-oriented sites (e.g., e-commerce) have lower revisits in the time range observed, meaning that revisit prediction should be adjusted by site.**



OTHER INTRA-SESSION MEASURES

- **Clickthrough rate (CTR): number of clicks (e.g., on an ad) divided by the number of times it was shown.**

- **Number of pages viewed (click depth): average number of contiguous pages viewed within a site**
  - Can be problematic if the website is ill-designed (too many clicks!!!)

- **Number of returns to the website within a session**
  - Useful for websites such as news aggregators, where returns indicate that the user believes there may be more information to glean from the site.

- **Conversion rate (mostly for e-commerce): fraction of sessions which end in a desired user action (e.g., purchase, download)**
  - Not all sessions are expected to result in a conversion, so this measure not always informative. However, it has the advantage of being closer to a website manager's goal.

INTER-SESSION ENGAGEMENT MEASURES

In general, these are the preferred measures of engagement

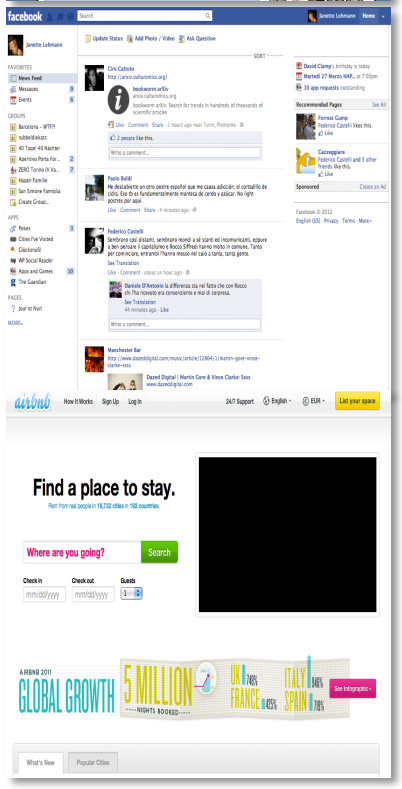
- **Direct value measurement:**
  - Lifetime value, as measured by ads clicked, monetization, etc.
- **Return-rate measurements:**
  - Fraction of return visits: How many users return for another visit?
  - Time between visits (inter-session time, absence time)
  - Number of distinct views (video)
- **Total use measurements:**
  - Total usage time per unit of time
  - Number of sessions per unit of time
  - Total view time per month (video)
  - Number of friends on site (social networks)

MODELS OF USER ENGAGEMENT BASED ON WEB ANALYTICS

... TOWARDS A TAXONOMY

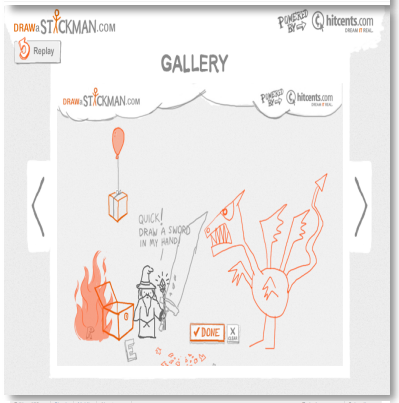
Games
Users spend much <u>time</u> per <u>visit</u>

Search
Users come <u>frequently</u> and do not stay long



Social media
Users come frequently and stay long

Service
Users visit site, when needed



Special
Users come on average once

News
Users come periodically

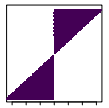
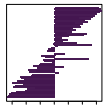
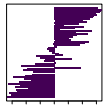
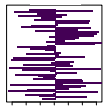
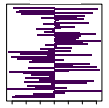
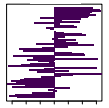
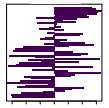
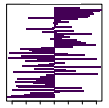
Online sites differ concerning their engagement!

DATA AND MEASURES

$$\tau_{intra} = 0.61$$

$$\tau_{inter} = 0.23$$

Interaction data, 2M users, July 2011, 80 US sites

Popularity	#Users	Number of distinct users	
	#Visits	Number of visits	
	#Clicks	Number of clicks	
Activity	ClickDepth	Average number of page views per visit.	
	DwellTimeA	Average time per visit $\tau = 0.33$	
Loyalty	ActiveDays	Number of days a user visited the site	
	ReturnRate	Number of times a user visited the site	
	DwellTimeL	Average time a user spend on the site.	

METHODOLOGY

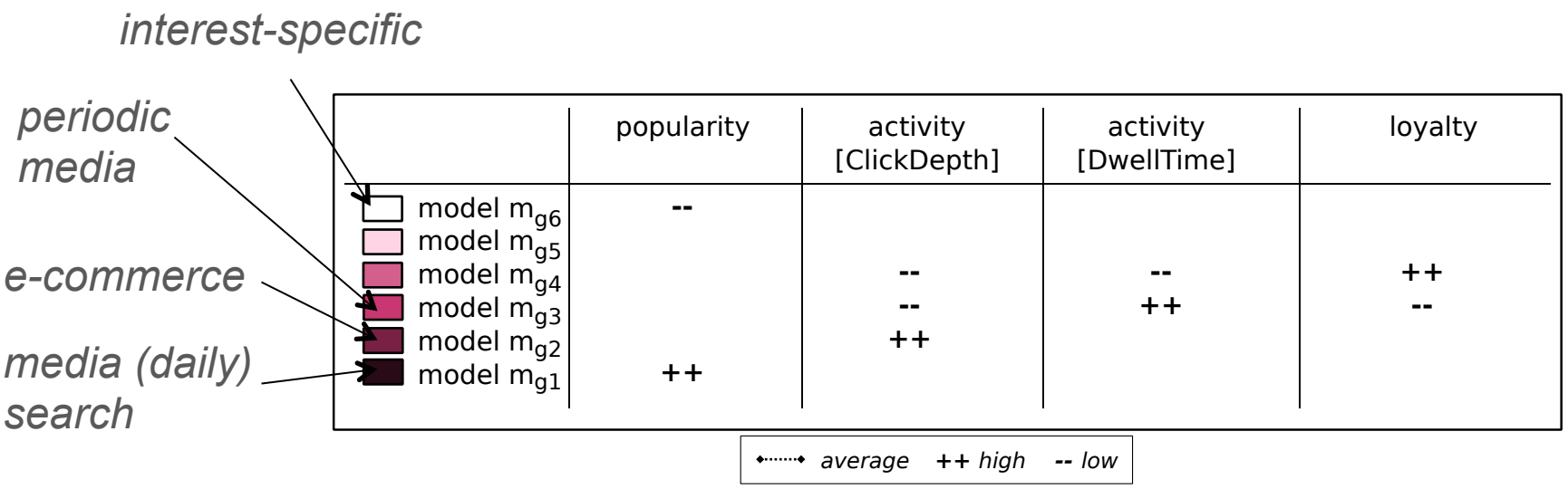
	General models	Time-based models
Dimensions	8 measures	weekdays, weekend 8 metrics per time span
#Dimensions	8	16
	Kernel k-means with *Kendall tau rank correlation kernel* *Nb of clusters based on eigenvalue distribution of kernel matrix* *Significant metric values with Kruskal-Wallis/Bonferonni*	
#Clusters (Models)	6	5
	Analysing cluster centroids = models	

MODELS OF USER ENGAGEMENT [6 GENERAL]

- **Popularity, activity and loyalty are <u>independent</u> from each other**
- **<u>Popularity</u> and <u>loyalty</u> are influenced by external and internal factors**

  e.g. frequency of publishing new information, events, personal interests

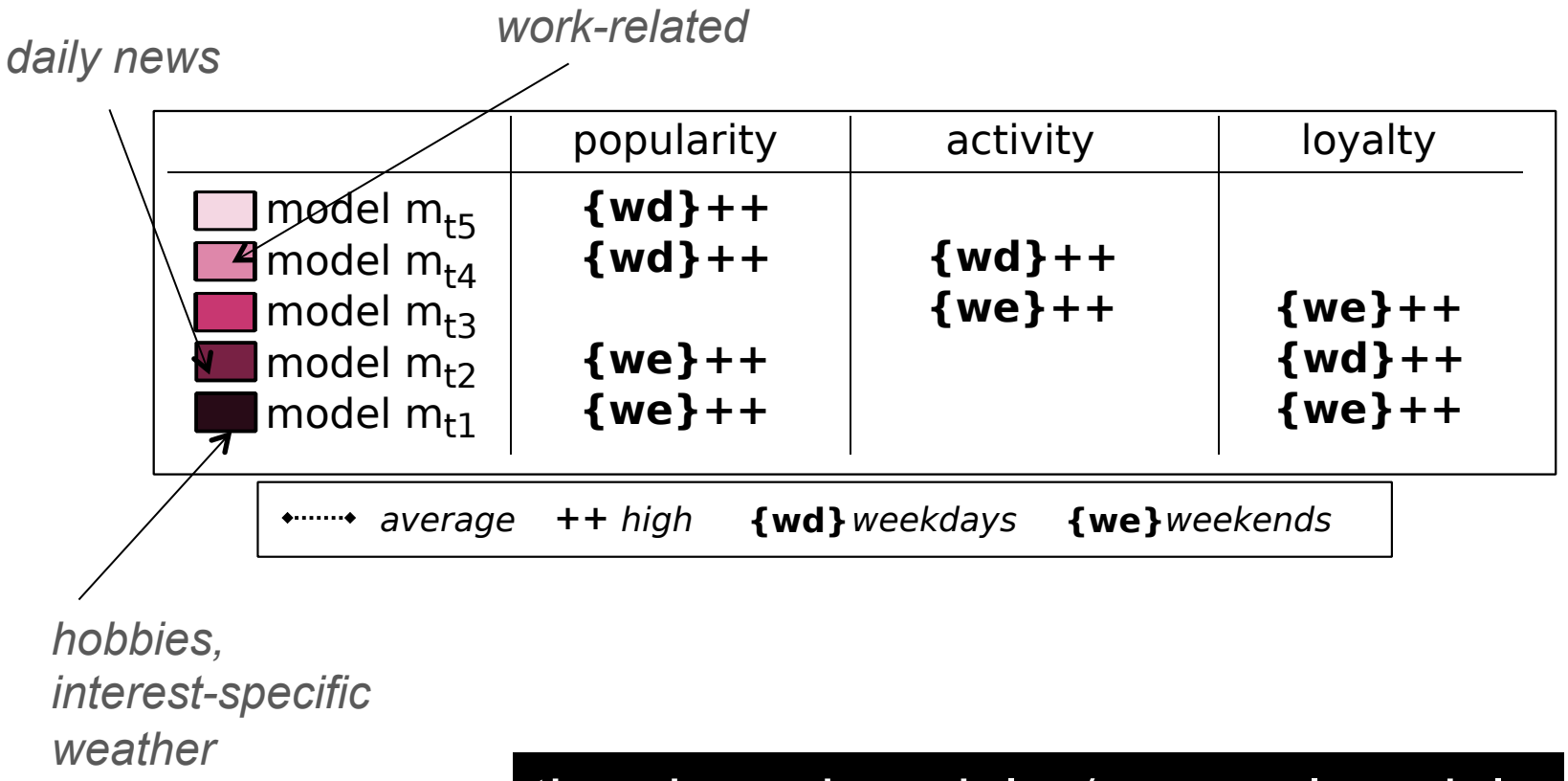
- **<u>Activity</u> depends on the structure of the site**



models based on engagement measures only

TIME-BASED [5 MODELS]

Models based on engagement over weekdays and weekend



time-based models ≠ general models

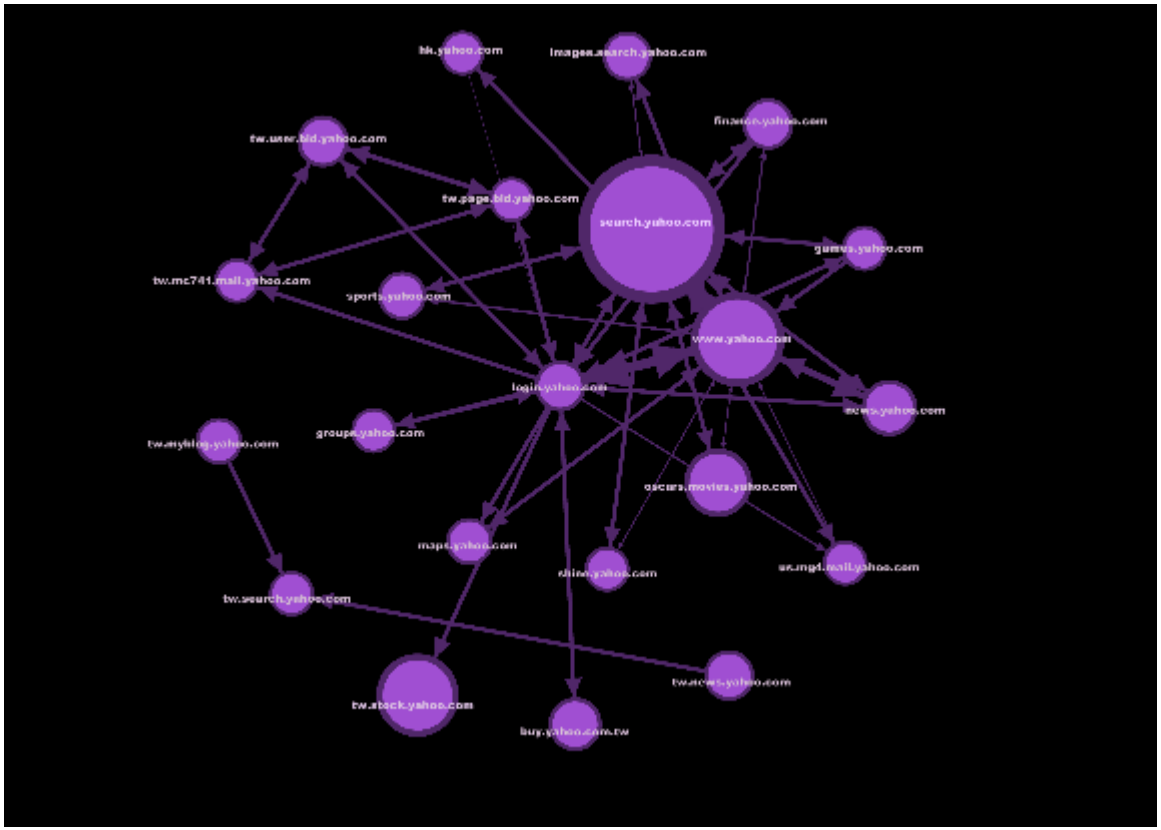
MODELS OF USER ENGAGEMENT

- **User engagement is complex and standard metrics capture only a part of it**
- **User engagement depends on time (and users)**
- **First step towards a taxonomy of models of user engagement … and associated measures**

- **What next?**
  - More sites, more models, more measures
  - User demographics, time of the day, geo-location, etc.
  - **Online multi-tasking**

(Lehmann et al, 2012)

ONLINE MULTI-TASKING



(fictitious navigation between sites within an online session)

181K users, 2 months browser data, 600 sites, 4.8M sessions

•only 40% of the sessions have no site **revisitation**

•hyperlinking, backpaging and teleporting

leaving a site is not a "bad thing!"

users spend more and more of their online session multi-tasking, e.g. emailing, reading news, searching for information → **ONLINE MULTI-TASKING**

navigating between sites, using browser tabs, bookmarks, etc

seamless integration of social networks platforms into many services

THE CONSULTANCY WORLD

(Peterson et al, September 2008)

- **Click Depth Index**: page views
- **Duration Index**: time spent
- **Recency Index**: rate at which users return over time
- **Loyalty Index**: level of long-term interaction the user has with the site or product (frequency)
- **Brand Index**: apparent awareness of the user of the brand, site, or product (search terms + social media)
- **Feedback Index**: qualitative information including propensity to solicit additional information or supply direct feedback
- **Interaction Index**: user interaction with site or product

$$E = \frac{1}{t}\left(\sum_{j=1}^{V} \frac{w_j(1+\%\Delta T_j) + \left(\sum_l L_l\right) + \left(\sum_f F_f\right) + \left(\sum_i I_i\right) + \left(\sum_b B_b\right)}{w_v} + \sum_{k=1}^{Z} \frac{dk}{dt}\right)$$

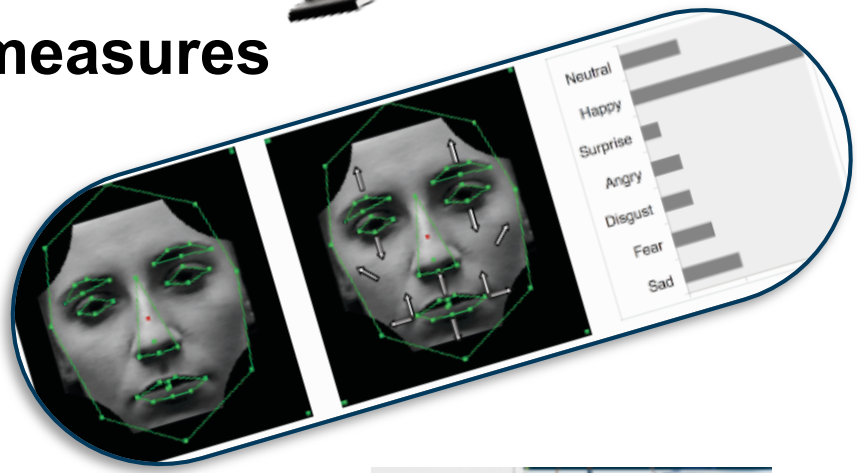
PART 1: FOUNDATIONS

APPROACHES BASED ON PHYSIOLOGICAL MEASURES

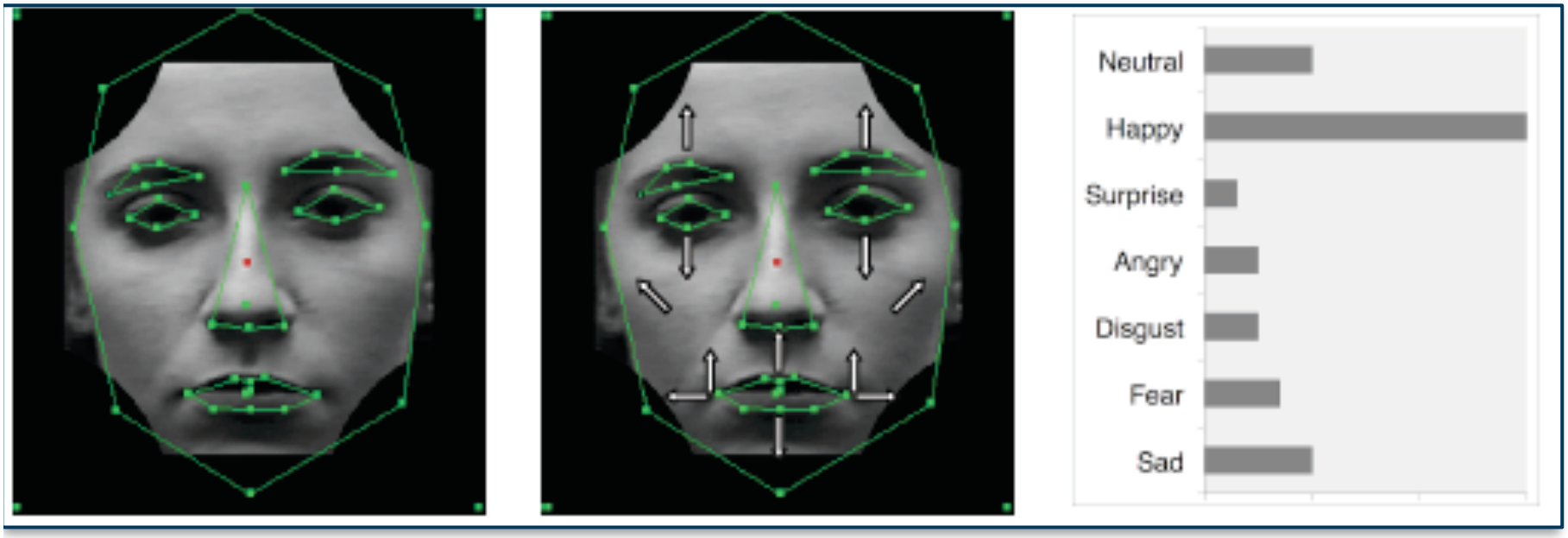
PHYSIOLOGICAL MEASURES

- **Eye tracking**
- **Mouse movement**
- **Face expression**
- **Psychophysiological measures**

  Respiration, Pulse rate
  Temperature, Brain wave,
  Skin conductance, …



FACIAL EXPRESSION AND SEARCH



16 subjects, facial expressions recorded while performing search tasks of various levels of difficultly.

learned model (based on support vector machine) shows that facial expressions provide good cues on topical relevance.

Potential application: personalised relevance feedback based on implicit cues.

(Arapakis et al, 2010)

WHAT IS PSYCHOPHYSIOLOGY?

- **The branch of physiology dealing with the relationship between physiological processes and thoughts, emotions, and behavior.**

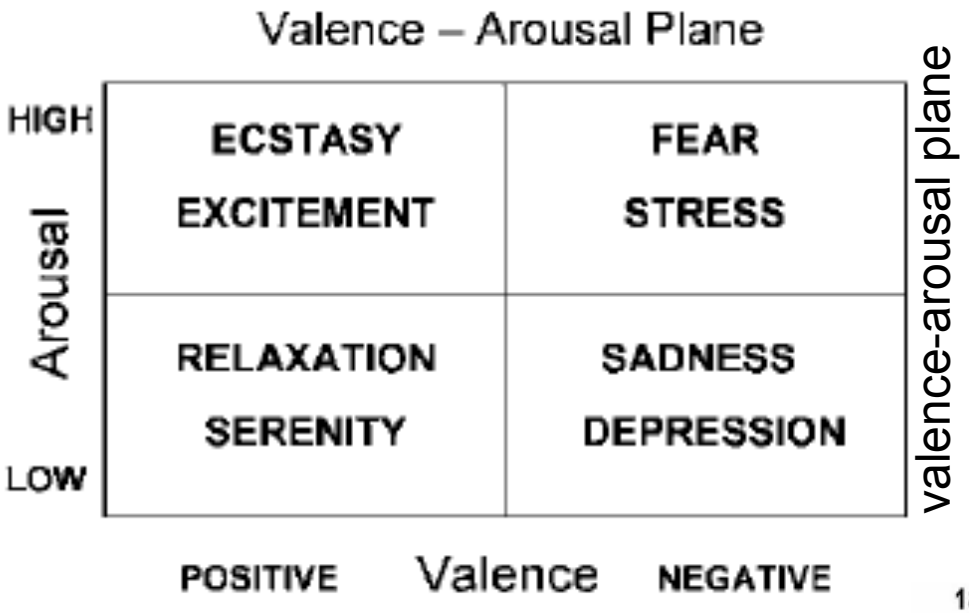
The body responds to psychological processes:

we exercise → we sweat

we get embarrassed → our cheeks get red and warm

- **Examples of measurements**
  - Electroencephalography (**EEG**) – measures the electrical activity of the brain through the scalp.
  - Cardiovascular measures – heart rate, **HR**; beats per minute, BPM; heart rate variability, **HRV**; vasomotor activity
  - Respiratory sensors – monitors oxygen intake and carbon dioxide output.
  - Electromyographic (**EMG**) sensors – measures electrical activity in muscles
  - Electrogastrogram (**EGG**) – measures changes in pupil diameter with thought and emotion (pupillometry) and eye movements
  - Galvanic skin response (**GSR**) sensors – monitors perspiration/sweat gland activity (also called Skin Conductance Level – **SCL**)
  - Temperature sensors – measures changes in blood flow and body temperature
  - Functional magnetic resonance imaging **(fMRI**) – measures brain activity by detecting associated changes in blood flow

FACEBOOK AND EMOTIONAL ENGAGEMENT (FLOW)



facebook

Lang model of emotions

SC = skin conductance
EMG = electromagnetic activity

relaxation (3mn, panorama pictures) → Facebook (3mn, free navigation) → stress (4mn, arithmetic tasks)

30 students

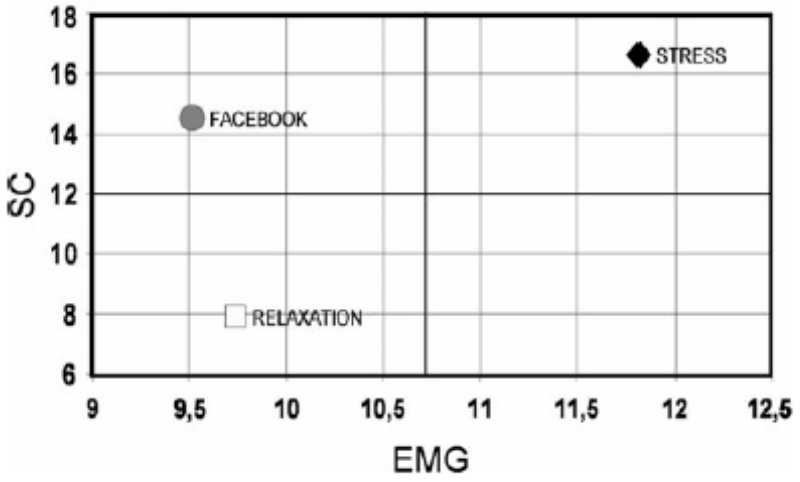


FIG. 9. The averaged data from all subjects for each epoch (relaxation, Facebook use, and stress) are plotted distributing the SC values along the Arousal axe (Y axe), and the EMG of CS along the Valence axe (X axe).

(Lang, 1995; Mauri et al, 2011)

PSYCHOPHYSIOLOGY – PROS AND CONS

Pros

- More objective data (not dependent on language, memory)
- Can be performed continuously during message/task processing
- Can provide information on emotional and attentional responses often not available to conscious awareness

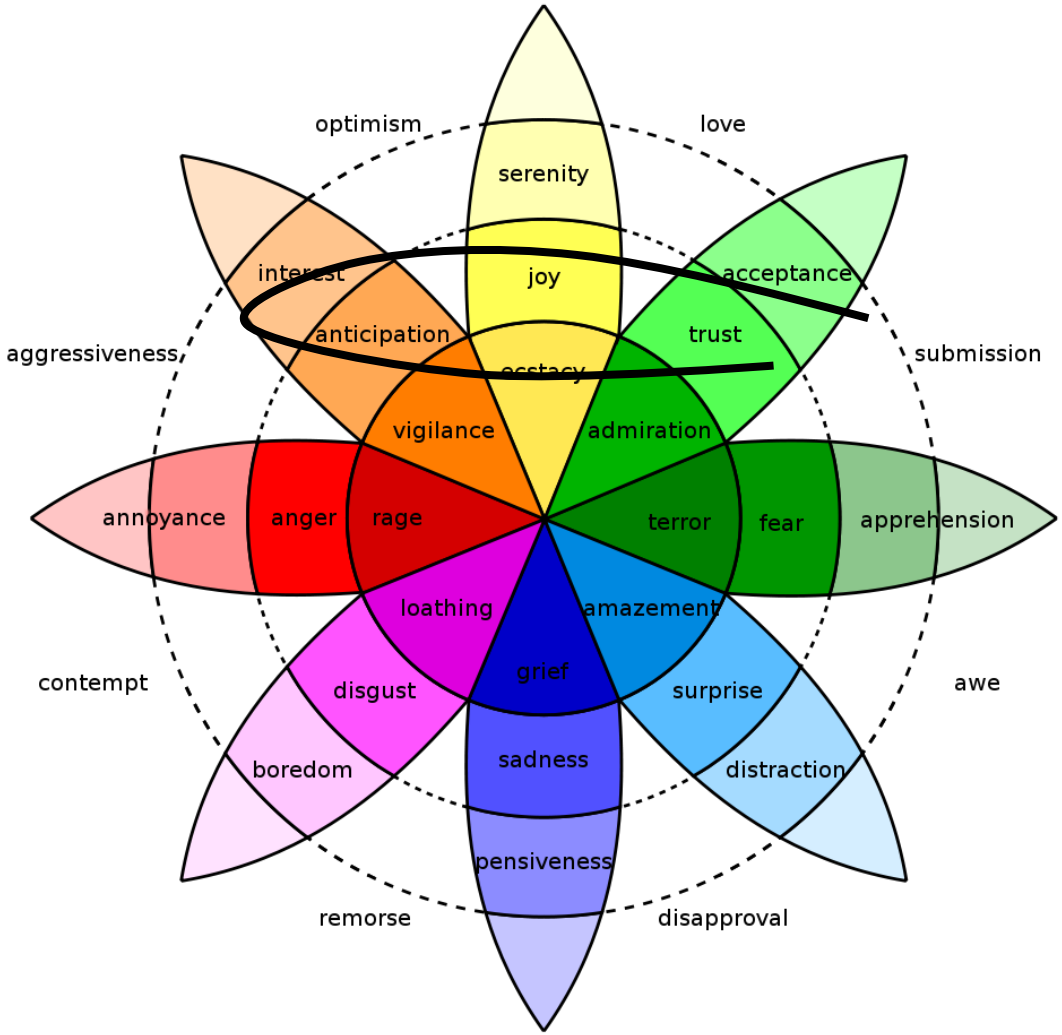
Cons

- Equipment expensive and can be cumbersome, and obtrusive
- Rarely a one-to-one correspondence between specific behaviors and physiological responses
- Difficult to operationalize and isolate a psychological construct
- Not applicable to large-scale

http://flavor.monell.org/~jlundstrom/research%20behavior.html

EMOTION, ENGAGEMENT AND MEASURES

Plutchik's emotion wheel



- **Anticipation:** Humans are curious.
- **Joy:** Happy users mean well engaged, repeat users.
- **Trust:** Users want to feel safe when interacting with your site.
- **More?**

http://uxdesign.smashingmagazine.com/2011/05/19/optimizing-emotional-engagement-in-web-design-through-metrics/

WHAT IS EYE TRACKING?

- **Process of measuring either the point of gaze (where one is looking) or the motion of an eye relative to the head.**

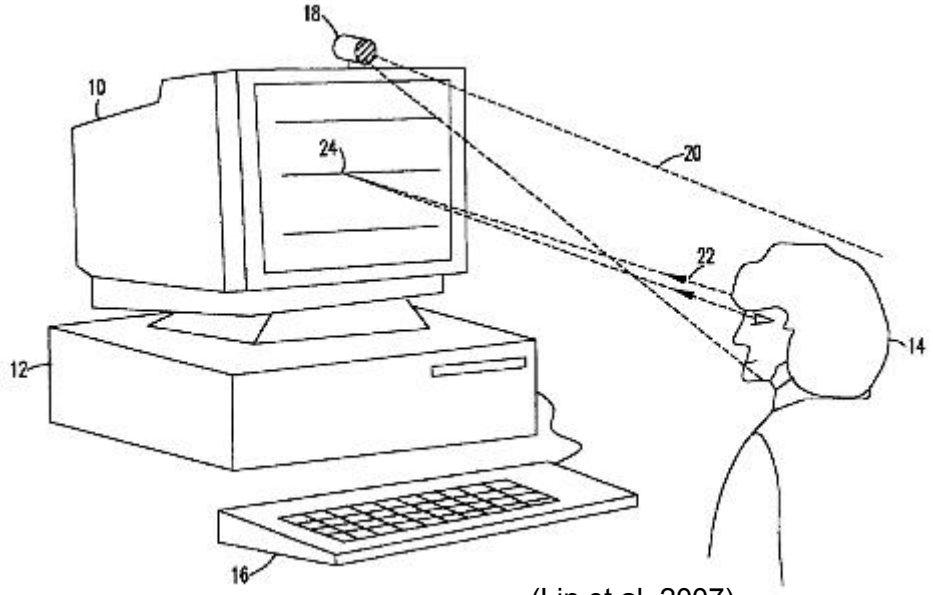
- **Eye tracker is a device for measuring eye positions and eye movement.**

- **Used in research on the visual system, in psychology, in cognitive linguistics and in product design.**

Examples of measures:

Time to First Fixation
Fixations Before
First Fixation Duration
Fixation Duration
Total Fixation Duration
Fixation Count
Visit Duration
Visit Count

whole screen or AOI (area of interest)



(Lin et al, 2007)

http://en.wikipedia.org/wiki/Eye_tracking

EYE TRACKING – ATTENTION AND SELECTION

Greece hopes for tourism rebound amid crisis

ATHENS, Greece - Many tourists see Athens as a launching pad for visiting the beaches and cute whitewashed buildings of the Greek islands. And the Aegean archipelago can be a great escape, especially during the nation's current economic crisis.

Thousands protest budget cuts at Calif. colleges

LONG BEACH, Calif. - More than 10,000 people marched, waved signs and occupied buildings at college campuses across California on Wednesday in a show of opposition to state budget cuts to education that could lead to ...

Colin Powell: Obama blew away the birthers

ORANGEBURG, S.C. - Colin Powell told graduates of South Carolina's premier historically black university that they were graduating during a tumultuous time that saw a royal wedding, a pope's beatification and a U.S. military assault that killed Osama bin Laden, "the worst person on earth."

Calif. woman shows off newly transplanted hand

LOS ANGELES - For the first time in five years, Emily Pennell has two hands.

Writer Harper Lee denies taking part in memoir

NEW YORK (AFP) - American writer Harper Lee, who rose to fame a half-century ago with her first and only novel, "To Kill A Mockingbird," denied Thursday that she had agreed to take part in a ...

Apple updates software to fix tracking glitch

SAN FRANCISCO (Reuters) - Apple Inc on Wednesday released a software update to fix a problem that enabled its mobile devices to collect and store customers' location data, making good on a promise it made last ...

Kvitova upsets Azarenka to take Madrid tennis title

MADRID (AFP) - Petra Kvitova defeated Victoria Azarenka 7-6 (7/3), 6-4 on Sunday to won the ATP-WTA Madrid Masters women's title and earn herself a place in the top 10 in the world rankings.

Obamas attend baptist church on Easter Sunday

WASHINGTON (AFP) - President Barack Obama and his family marked Easter Sunday by attending a service at an African-American baptist church in Washington, standing to clap the 120-strong choir.

18 users, 16 tasks each
(chose one story and rate it)

eye movement recorded

Attention (gaze)
interest has no role
position > saliency

Selection
mainly driven by interest
position > attention

(Navalpakkam et al., 2012)

EYE TRACKING – PROS AND CONS

- **Pros**
  - Lots of details (fine-grained data/resolution)
  - Offers direct measure of user attention + what they are looking at
  - Offers insights into how people consume & browse web pages + why they fail at clicking on something
- **Cons**
  - Not scalable
  - Slow and expensive
  - Not natural environment (e.g. at home)
    - Behavior IS different in a lab setting

Can mouse movement act as a (weak) proxy of gaze?

WHAT IS MOUSE TRACKING? (also known as cursor tracking)

- **Using software (JavaScript) to collect user mouse cursor positions on computer/web interface**
- **Aim to provide information about what people are doing, typically to improve the design of an interface**
- **How does gaze is measured by an eye tracker relates to mouse movement as recorded**
- **Studies and applications**
  - Attention on web pages
  - Relevance of search results
    - As a proxy of relevance
    - As additional and complimentary signal

MOUSE VS GAZE – ATTENTION ON WEB PAGES

- **90 users on 6 Yahoo! Finance articles – rich media content**
- **3 treatments:**
  - ad always on top; ad top right + random; random (6 positions)
- **Reading tasks + post-questionnaires**

ad avoidance

similar patterns
shift of attention from top-left to right as ad position change

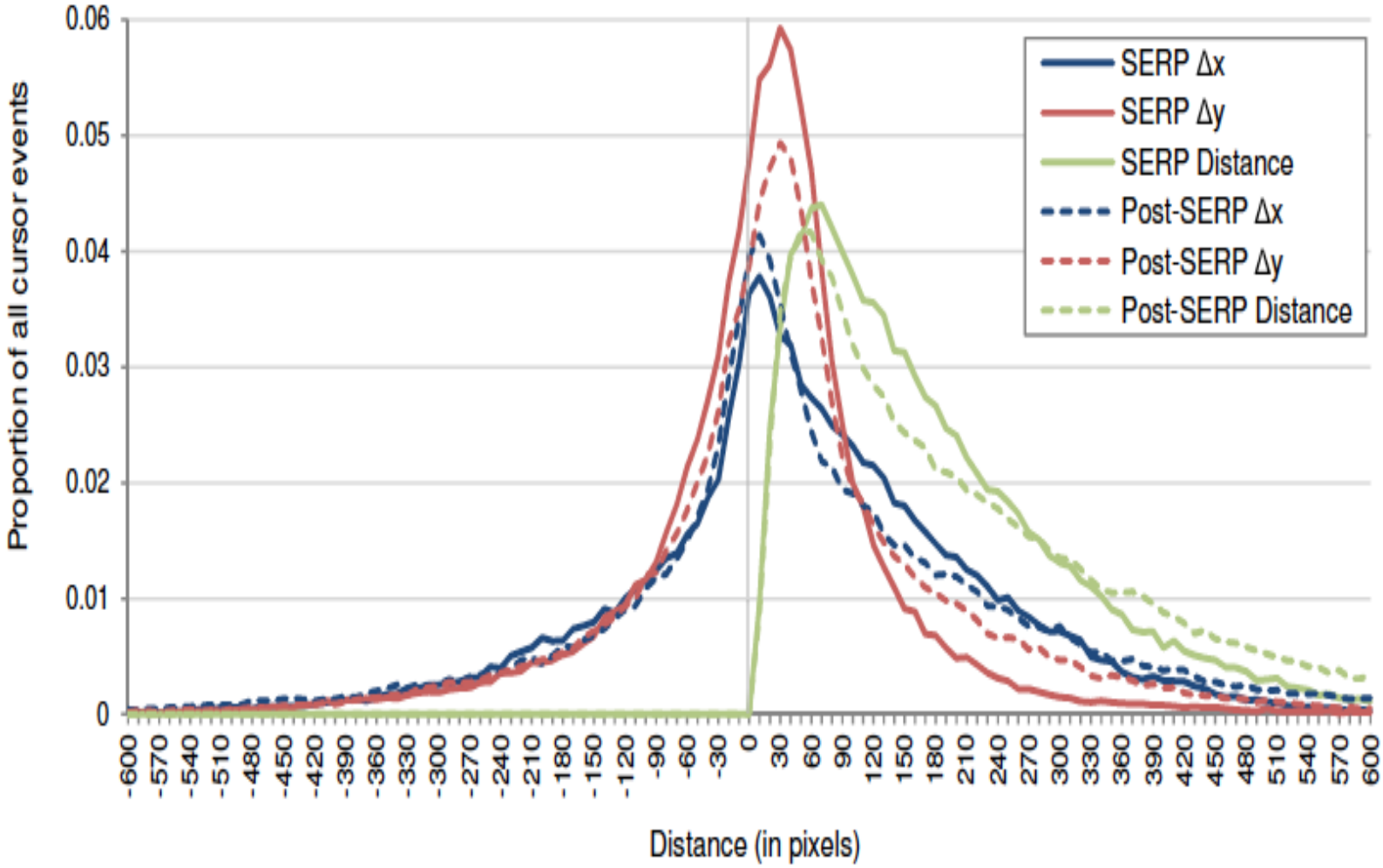
similar patterns
more at top position
and longer dwell
left better than right

similar patterns
visit ad sooner & more time to process content when ad position moves

Similar patterns between gaze and mouse in terms of user attention when manipulating conditions (here ads)
Interesting results for “ads”

(Navalpakkam & Churchill, 2012)

GAZE VS MOUSE - DISTANCE



(Huang, White & Dumais, 2011)

GAZE VS CURSOR - FACTORS

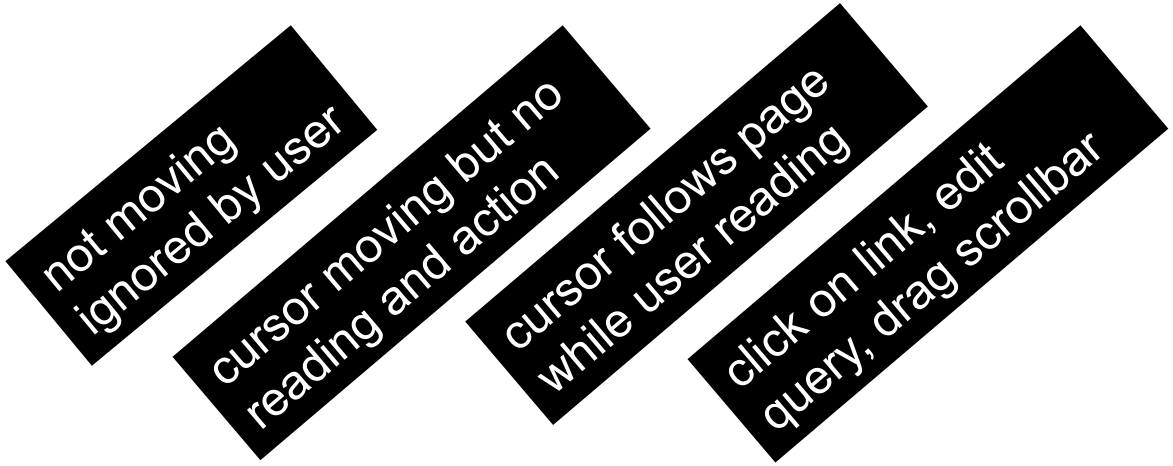
(Huang et al, 2012)

- **38 users and 32 search tasks (navigational + informational)**
- **Age or gender does not seem to be a factor**
- **Task does not seem to be a factor (others found the opposite)**

  (using click entropy to classify a a query)
- **User individual behavior seem to matter more**
- **Gaze leads the cursor**
- **Stronger alignment when search result page loads**
- **Cursor behaviors: alignment increases**

inactive < examining < reading < action < click

58.8% 31.9% 2.5% 5.7%



classification (heuristic-based) informed by watching replay of user interactions

CLICK VS CURSOR – HEATMAP

(Huang et al, 2011)

the role of hovering?

- **Estimate search result relevance**

(Bing - Microsoft employees – 366,473 queries; 21,936 unique cookies; 7,500,429 cursor move or click)

Click positions

Cursor movement positions

Figure 2. Heatmaps of all click positions (left) and recorded cursor positions (right) for the query *[lost finale explanation]*. Heavy interaction occurs in red/orange/yellow areas, moderate interaction in green areas, light interaction in blue areas.

MOUSE MOVEMENT – WHAT CAN HOVERING TELL ABOUT RELEVANCE?

Table 3. Correlations between click and hover features and relevance judgments for queries with and without clicks.

Result clicks or no clicks	Feature source	Correlation with human relevance judgments
Clicks (N=1194)	Clickthrough rate (c)	0.42
	Hover rate (h)	0.46
	Unclicked hovers (u)	-0.26
	Max hover time (d)	-0.15
	Combined[1]	**0.49**
No clicks (N=96)	Hover rate	0.23
	Unclicked hovers	0.06
	Max hover time	0.17
	Combined[2]	**0.28**

Cickthrough rate:
% of clicks when URL Shown (per query)

Hover rate:
% hover over URL (per query)

Unclicked hover:
Media time user hovers over URL but no click (per query)

Max hover time:
Maximum time user hover over a result (per SERP)

(Huang et al, 2011)

MOUSE MOVEMENT – WHAT CAN HOVERING TELL ABOUT ABANDONMENT?

(Huang et al, 2011)

- **Abandonment (a engagement metric in search) is when there is no click on the search result page**
  - User is dissatisfied (bad abandonment)
  - User found result(s) on the search result page (good abandonment)
- **858 queries (21% good vs. 79% abandonment manually examined)**
- **Cursor trail length**
  - Total distance (pixel) traveled by cursor on SERP
  - Shorter for good abandonment
- **Movement time**
  - Total time (second) cursor moved on SERP
  - Slower when answers in snippet (good abandonment)
- **Cursor speed**
  - Average cursor speed (pixel/second)
  - Slower when answers in snippet (good abandonment)

RELEVANCE & DWELL TIME

… we recall that in search

- **Dwell time – on landing page (post search result)**
  - Although a good indicator of user interest/relevance, not reliable on its own
  - Time spending reading a document (result) has shown to improve search quality
    - Short dwell time a good indication of non-relevance
  - BUT
    - Interpreting long dwell-time not so straight-forward

RELEVANCE & CURSOR

(Guo & Agichtein, 2012)

(a) relevant (dwell time: 30s)

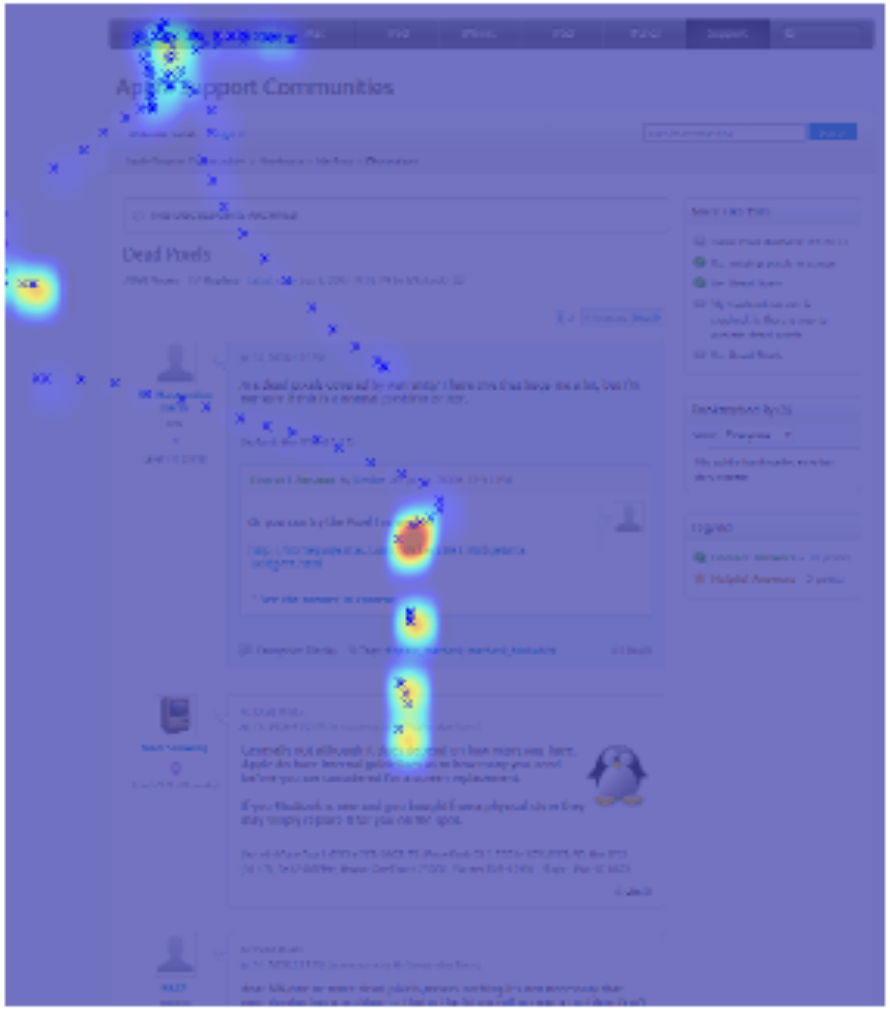
(b) non-relevant (dwell time: 30s)

"reading" cursor heatmap of relevant document vs "scanning" cursor heatmap of non-relevant document (both dwell time of 30s)

RELEVANCE & CURSOR

(Guo & Agichtein, 2012)

(a) relevant (dwell time: 70s)



(b) non-relevant (dwell time: 80s)

“reading” a relevant long document vs “scanning” a long non-relevant document

OUTLINE

- **Introduction and Scope**

- **Part I - Foundations**
  1. Approaches based on self-report measures
  2. Approaches based on web analytics
  3. Approaches based on physiological measures

- **Part II – Advanced Aspects**
  1. Measuring user engagement in mobile information searching
  2. Networked user engagement
  3. Combining different approaches

- **Conclusions**
- **Bibliography**

PART 2: ADVANCED ASPECTS

MOBILE INFORMATION SEEKING

MOBILE USER ENGAGEMENT

- Mobile devices are changing the ways in which we are learning, working, and communicating.
- The role of device has not been considered in (published) studies of user engagement.
- However … related work has been done in the UX literature.

DIARY STUDIES

1. **Komaki et al, 2012**
   - Context heavily influenced search behavior

2. **Nylander et al, 2009**
   - General preference for using mobile, even when an alternative was available (51% of instances)
   - Mobile use influenced by: technical ease and functionality, and convenience, laziness, and integration with social life and daily activities

3. **Church & Smythe, 2009; Church & Oliver, 2011**
   - Emphasized location and time as key factors in mobile use

FIELD STUDIES

- **Oulasvirta et al, 2005**
  - Attention shifting between the mobile device and the external environment

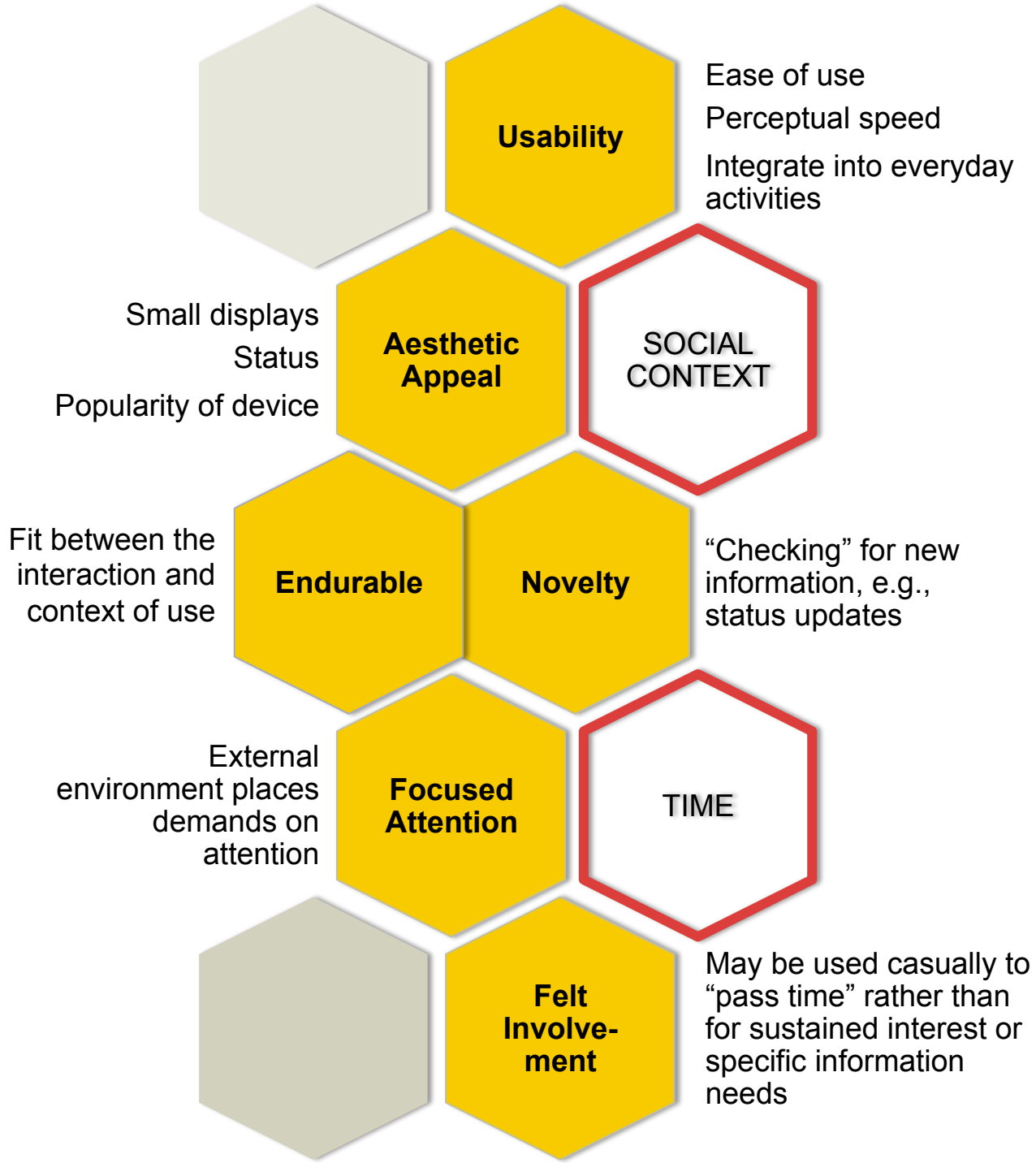
- **Gökera & Myrhaugb, 2008**
  - Context closely tied to perceived relevance and value of information

- **Battarbee & Koskinen, 2005**
  - Emotional response of information sharing and communication with friends in everyday life

BUILDING A MODEL OF ENGAGEMENT BASED ON UX LITERATURE

- **User experience (UX) literature suggests that:**
  - Users must focus attention on the mobile task *and* the external environment (Oulasvirta et al., 2005).
  - 63% of mobile searches were social in nature (Teevan et al. 2011).
  - Mobile devices with constant connectivity are often 'habit-forming' (Oulasvirta et al., 2012)
  - Time motivates mobile phone use (Tojib & Tsarenko, 2012).

Therefore …


Usability
Ease of use
Perceptual speed
Integrate into everyday activities
Small displays
Status
Popularity of device
Aesthetic Appeal
SOCIAL CONTEXT
Fit between the interaction and context of use
Endurable
Novelty
"Checking" for new information, e.g., status updates
External environment places demands on attention
Focused Attention
TIME
Felt Involve-ment
May be used casually to "pass time" rather than for sustained interest or specific information needs

STUDYING MOBILE USER ENGAGEMENT (IN PROGRESS)

While conversing about their carbon footprints, Mary and John could not decide which of their cars are more energy-efficient

Point of Engagement

Period of Sustained Engagement

"Let's look it up!"

Disengagement

"We're late for class. Better go!"

Re-engagement

Upon returning home, John decides to search more about hybrid cars on his computer

(Absar, O'Brien, Halbert & Trumble, 2013)

MOBILE USER ENGAGEMENT: EXPLORATORY STUDY METHODS

Interview 1 → **Mobile Diary Collection Period** → Interview 2

- Demographics
- Mobile search behaviours
- User Engagement
- Instructions for study

TEXT + PHOTO

(Photovoice, Wang & Burris, 1997)

- Selection of Five Diary Entries
- Semi-structured interview
- General questions of mobile use

ENGAGEMENT WITH MOBILE APPS

- **Focused on branded mobile apps, interactive marketing tools**

- **Methodology: identification and analysis of branded apps**
  - 2010 Interbrand Top 100 Global Brands + iTunes app store
  - Analysis of *features* and *content* on the branded app according to: vividness, novelty, motivation, control, customization, feedback, and multiplatforming
  - Distinguished product and service branded apps

- **Almost all apps incorporated at least one of the seven engagement attributes:**
  - control (97.2%), customization (85.8%), vividness (78.3%: entire app, 86.8%: entry page), multiplatforming (70.8%), motivation (62.3%), feedback (55.7%), and novelty (11.3%).

(Kim, Lin & Sung, 2013)

PART 2: ADVANCED ASPECTS

NETWORKED USER ENGAGEMENT

DOWNSTREAM ENGAGEMENT

or website

NO MAN IS AN ISLAND, ENTIRE OF ITSELF

- **Basic premises:**
  - The success of a website depends not only on itself, but also on its environment.
  - This is particularly relevant for companies running networks of properties or services

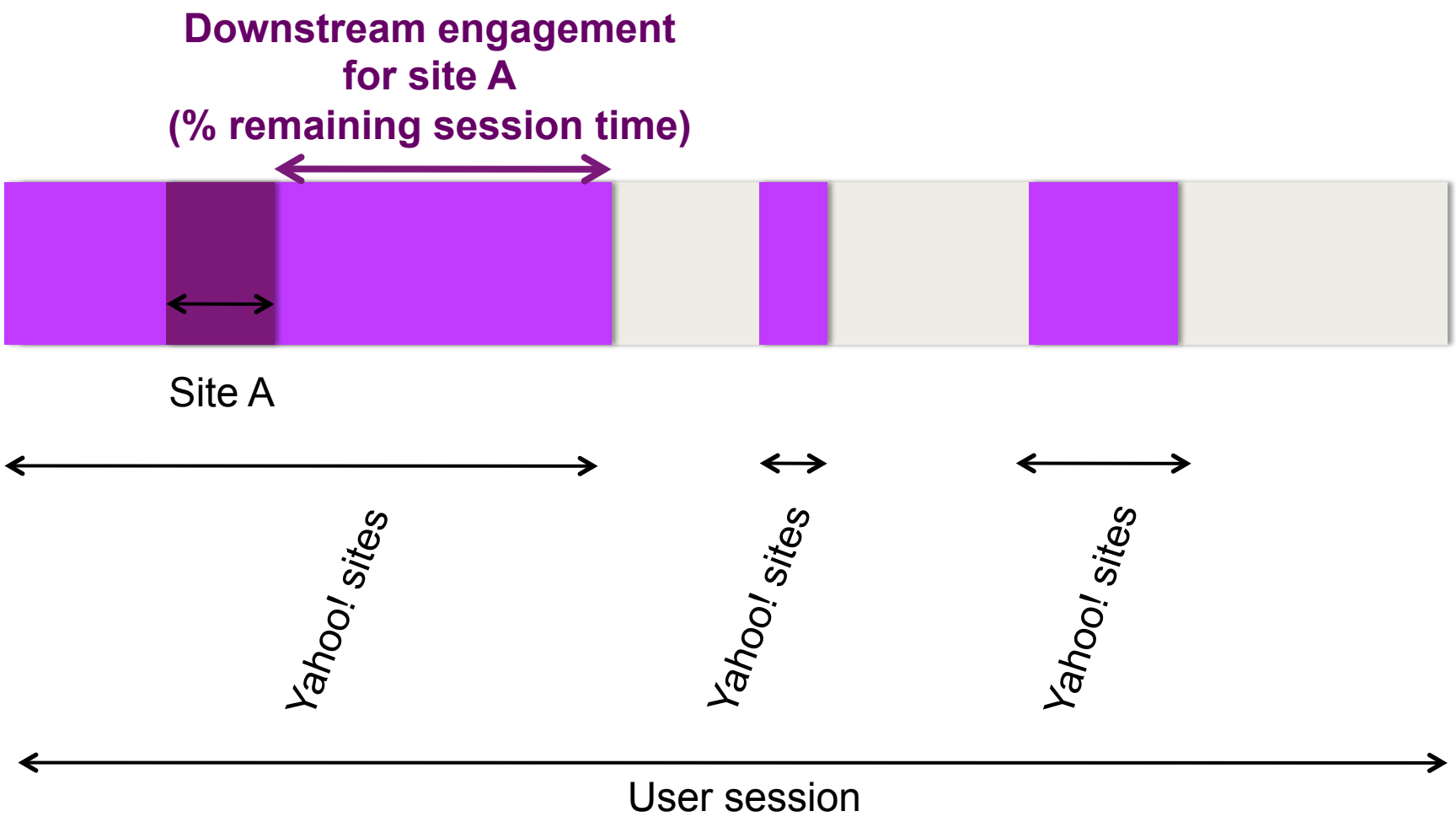
USER BEHAVIOR WITHIN A NETWORK OF SITES

NETWORKED USER ENGAGEMENT:

ENGAGEMENT ACROSS A NETWORK OF SITES

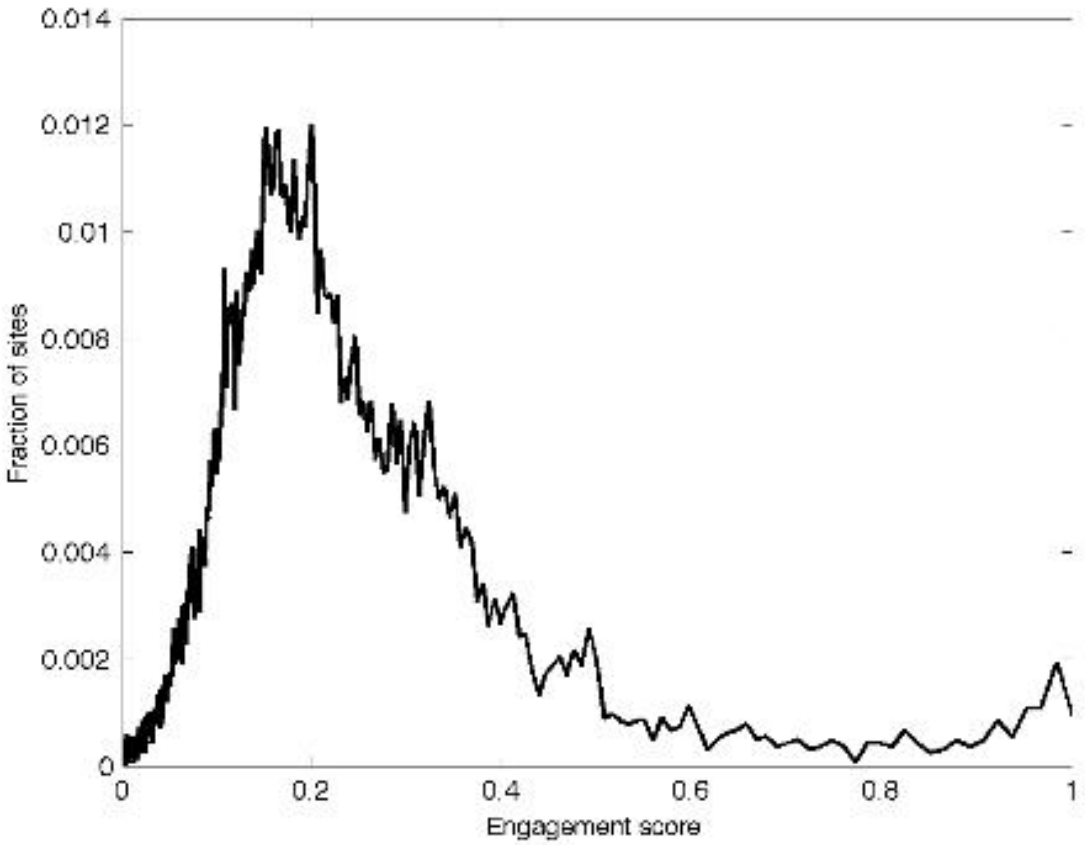
- Large online providers (AOL, Google, Yahoo!, etc.) offer not one service (site), but a network of sites
- Each service is usually optimized individually, with some effort to direct users between them
- Success of a service depends on itself, but also on how it is reached from other services (user traffic)
- Users switch between sites within an online session, several sites are visited and the same site is visited several times (online multi-tasking)

MEASURING DOWNSTREAM ENGAGEMENT



DISTRIBUTION OF DOWNSTREAM ENGAGEMENT SCORES

- **Downstream engagement is not highly correlated with intra-site measures of engagement such as dwell time ($\varrho = -0.05$, $p < 10^{-5}$).**
- **Downstream engagement is negatively correlated with inter-session measures such as revisits ($\varrho = -0.26$, $p < 10^{-5}$).**

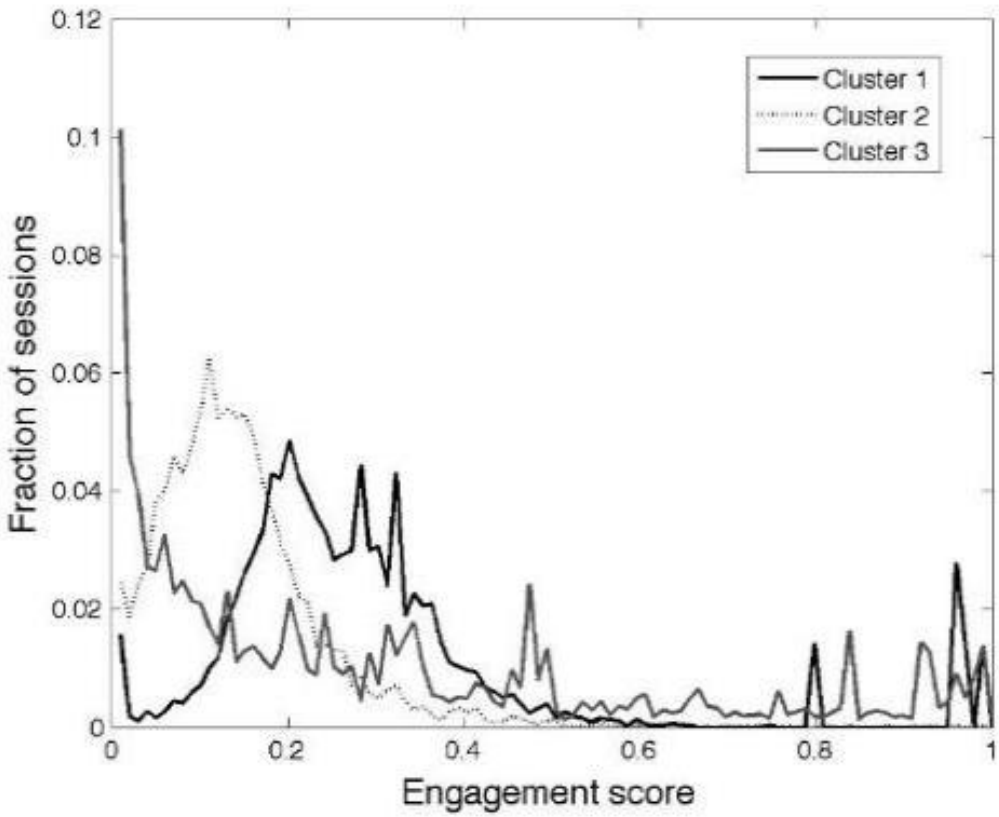


(19.4M sessions, 265,000 users, 50 sites)

CLUSTERED DISTRIBUTION OF DOWNSTREAM ENGAGEMENT SCORES

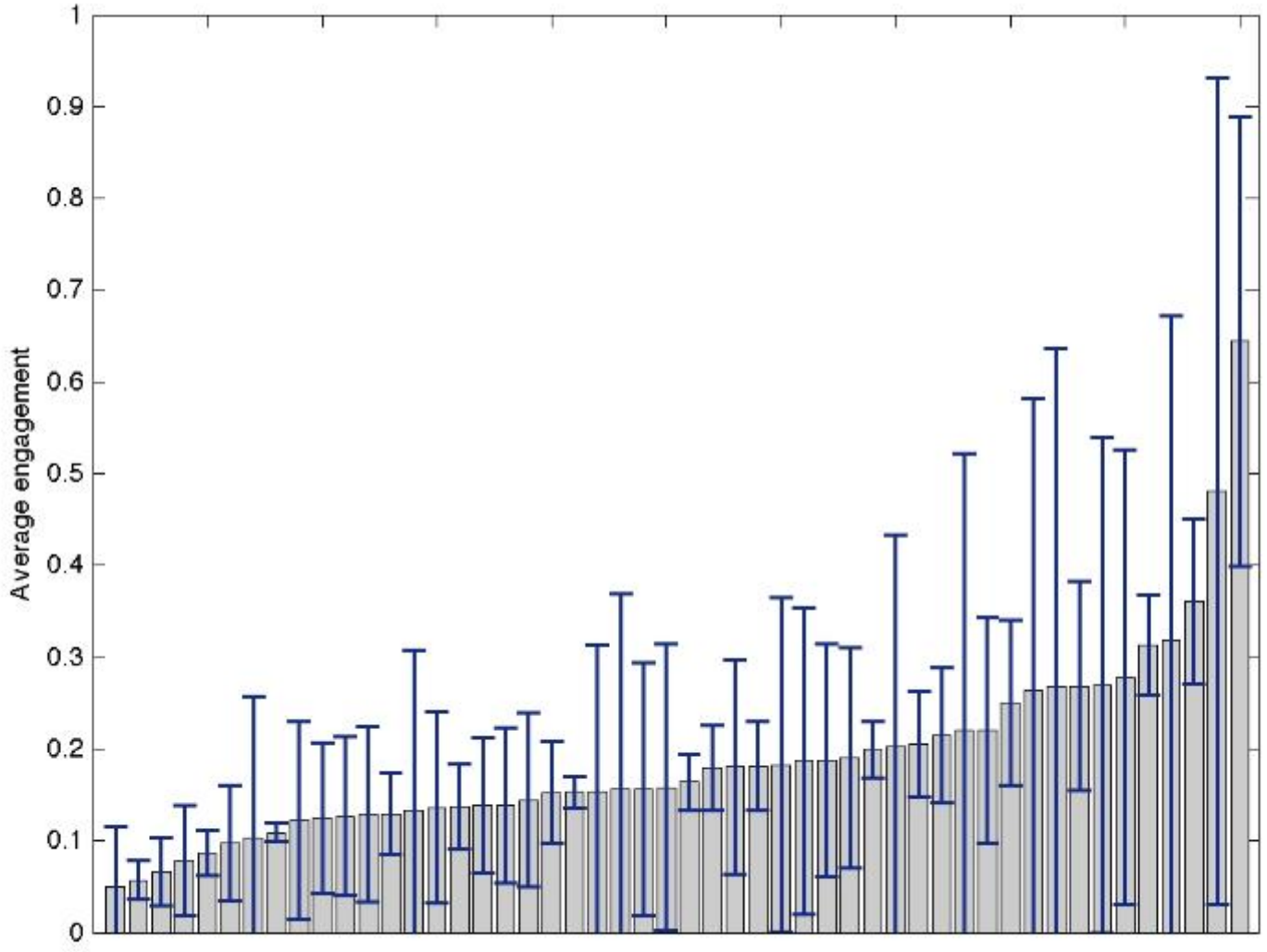
There are different modes of downstream engagement according to site type.

There are no obvious characteristics of websites that would indicate their downstream distribution.



(19.4M sessions, 265,000 users, 50 sites)

DISTRIBUTION OF DOWNSTREAM ENGAGEMENT TO A LIST OF YAHOO! WEBSITES



Varies across and within websites (19.4M sessions, 265,000 users, 50 sites)

NETWORKED USER ENGAGEMENT

- **Downstream engagement**
  - Varies significantly across sites
  - Exhibits different distributions according to site type (but we don't understand enough!)

- **Other measures of networked user engagement?**

- **Applications to companies with several services but also to increasing "tightly" connected services (news and social media)**

- **Let us not forget increased online multitasking**

- **Next**: Can we quantify the network effect?

(Yom-Tov et al., 2013)

PART 2: ADVANCED ASPECTS

COMBINATIONS OF APPROACHES

MEASURING USER ENGAGEMENT – WE RECALL

	Measures	Characteristics
Self-reported engagement	Questionnaire, interview, report, product reaction cards, think-aloud	Subjective Short- and long-term Lab and field Small-scale Product outcome
Cognitive engagement	Task-based methods (time spent, follow-on task) Neurological measures (e.g. EEG) Physiological measures (e.g. eye tracking, mouse-tracking)	Objective Short-term Lab and field Small-scale and large-scale Process outcome
Interaction engagement	Web analytics + "data science" *metrics + models*	Objective Short- and long-term Field Large-scale Process

COMBINATION OF APPROACHES

SEVERAL STUDIES

USER ENGAGEMENT

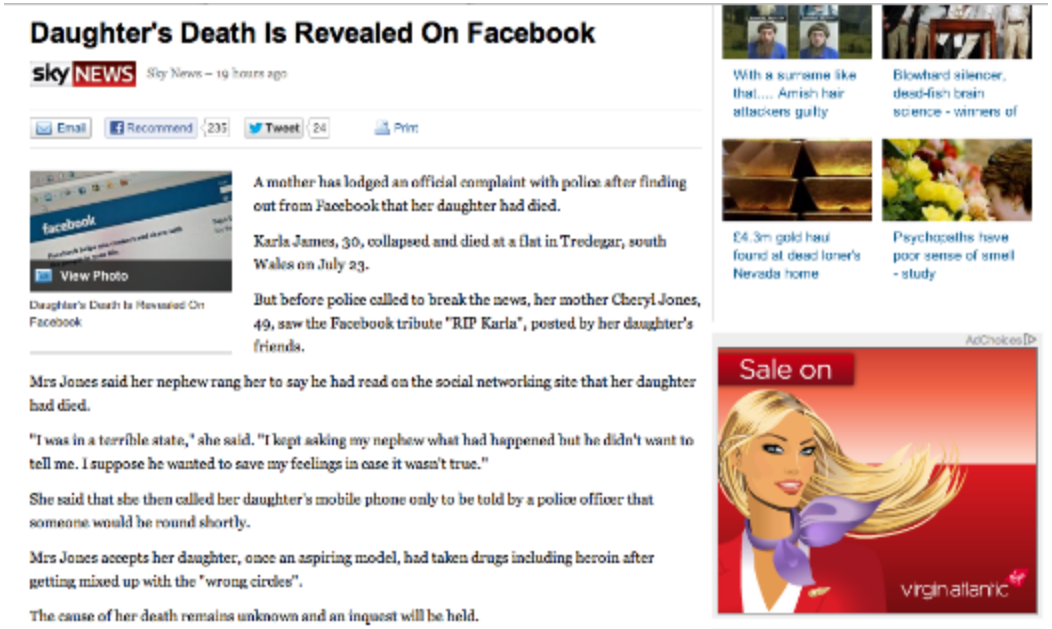
interaction engagement

self-reported engagement

cognitive engagement

STUDY I: GAZE AND SELF-REPORTING

- **News + comments**
- **Sentiment, interest**
- **57 users (lab-based)**
- **Reading task (114)**



- **Questionnaire (qualitative data)**
- **Record mouse tracking, eye tracking, facial expression, EEG signal (quantitative data)**

Three metrics: gaze, focus attention and positive affect

INTERESTING CONTENT PROMOTE USERS ENGAGEMENT METRICS

- **All three metrics:**
  - focus attention, positive affect & gaze
- **What is the right trade-off?**
  - news is news ☺
- **Can we predict?**
  - provider, editor, writer, category, genre, visual aids, …, **sentimentality**, …
- **Role of user-generated content (comments)**
  - As measure of engagement?
  - **To promote engagement?**

LOTS OF SENTIMENTS BUT WITH NEGATIVE CONNOTATIONS!

- **Positive affect (and interest, enjoyment and wanted to know more) correlates**
  - Positively (🡅) with sentimentality (lots of emotions)
  - Negatively (🡇) with positive polarity (happy news)

SentiStrenght (from -5 to 5 per word)

sentimentality: sum of absolute values (amount of sentiments)
polairity: sum of values (direction of the sentiments: positive vs negative)

(Thelwall, Buckley & Paltoglou, 2012)

EFFECT OF COMMENTS ON USER ENGAGEMENT

- **6 ranking of comments:**
  - **most replied, most popular, newest**
  - sentimentality high, sentimentality low
  - polarity plus, polarity minus
- **Longer gaze on**
  - newest and most popular for interesting news
  - most replied and high sentimentality for non-interesting news
- **Can we leverage this to prolong user attention?**

GAZE, SENTIMENTALITY, INTEREST

- **Interesting and "attractive" content!**
- **Sentiment as a proxy of focus attention, positive affect and gaze?**

- **Next**
  - Larger-scale study
  - Other domains (beyond daily news!)
  - Role of social signals (e.g. Facebook, Twitter)
  - Lots more data: mouse tracking, EEG, facial expression

(Arapakis et al., 2013)

STUDY II: MOUSE TRACKING AND SELF-REPORTING

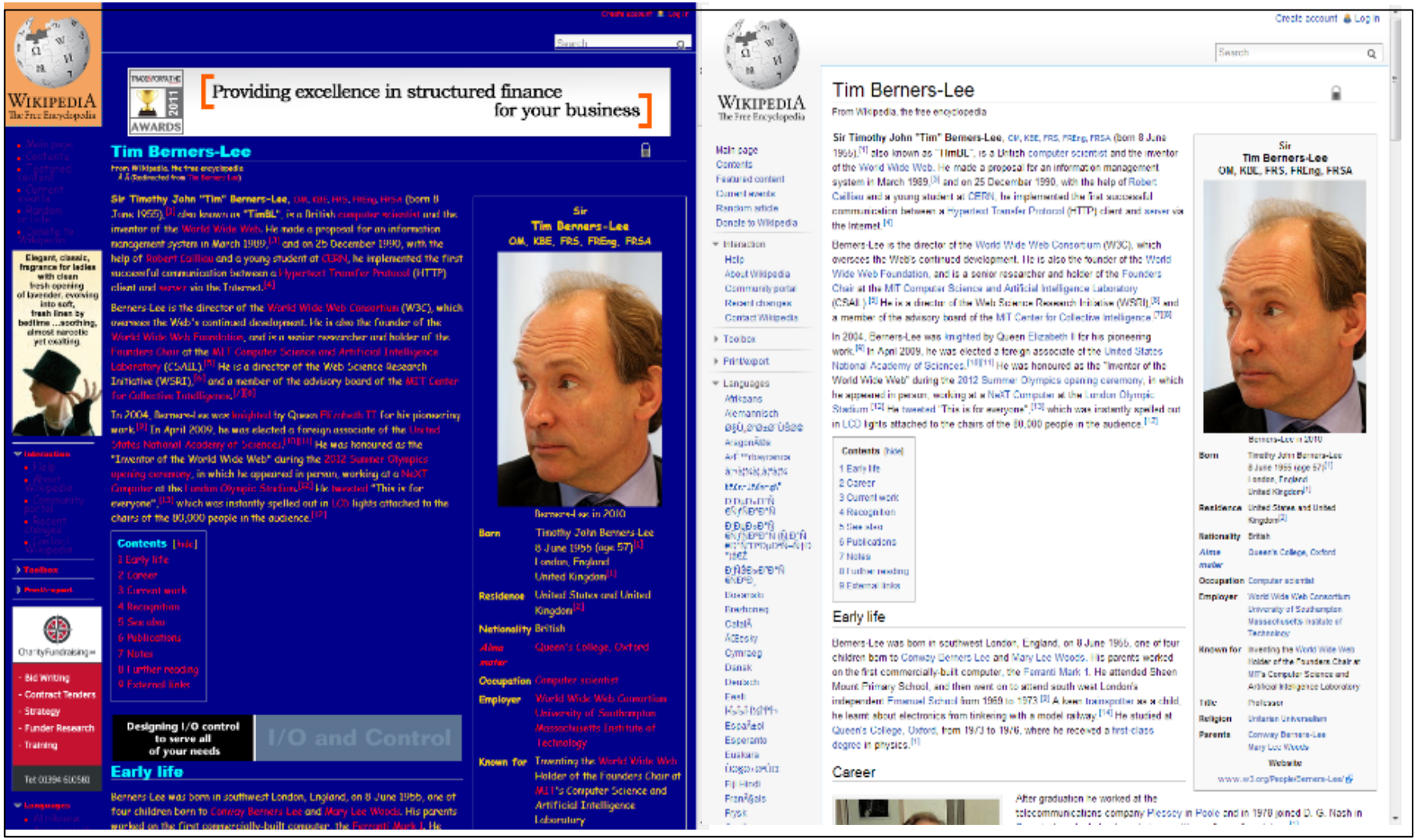
- **324 users from Amazon Mechanical Turk** (between subject design)
- Two domains (BBC News and Wikipedia)
- Two tasks (reading and search)
- "Normal vs Ugly" interface

- **Questionnaires (qualitative data)**
  - focus attention, positive effect, novelty,
  - interest, usability, aesthetics
  - + demographics, handeness & hardware

- **Mouse tracking (quantitative data)**
  - movement speed, movement rate, click rate, pause length, percentage of time still

“Ugly” vs “Normal” Interface (BBC News)

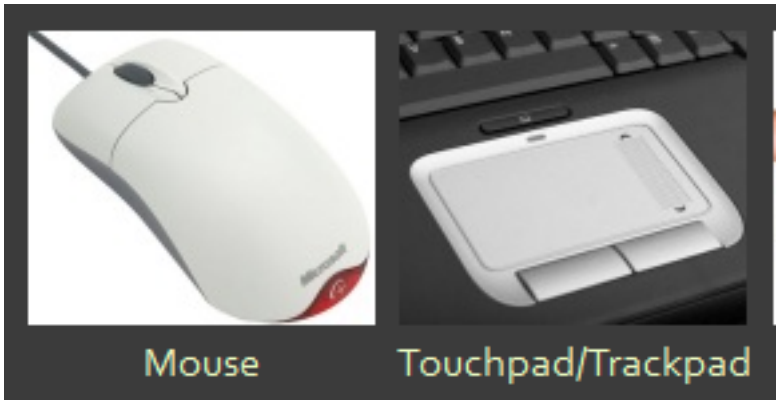
“Ugly” vs “Normal” (Wikipedia)



MOUSE TRACKING CAN TELL ABOUT

- **Age**

- **Hardware**
  - Mouse
  - Trackpad



- **Task**
  - Searching: *There are many different types of phobia. What is Gephyrophobia a fear of?*
  - Reading: (Wikipedia) *Archimedes, Section 1: Biography*

MOUSE TRACKING COULD NOT TELL MUCH ON

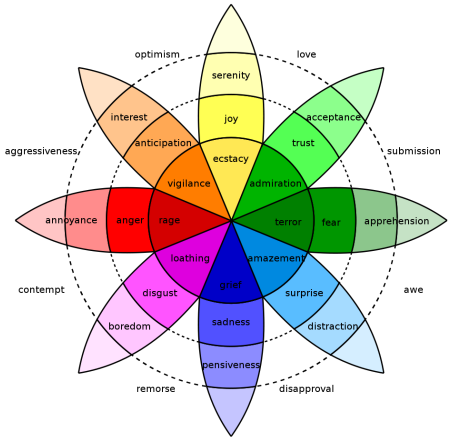
- **focused attention and positive affect**
- **user interests in the task/topic**

- **BUT BUT BUT BUT**
  - "ugly" variant did not result in lower aesthetics scores
  - although BBC > Wikipedia

- **BUT – the comments left …**
  - Wikipedia: "*The website was simply awful. Ads flashing everywhere, poor text colors on a dark blue background.*"; "*The webpage was entirely blue. I don't know if it was supposed to be like that, but it definitely detracted from the browsing experience.*"
  - BBC News: "*The website's layout and color scheme were a bitch to navigate and read.*"; "*Comic sans is a horrible font.*"

MOUSE TRACKING AND USER ENGAGEMENT

- **Task and hardware**
- **Do we have a Hawthorne Effect???**
- **“Usability” vs engagement**
  - “Even uglier” interface?
- **Within- vs between-subject design?**



- **What next?**
  - Sequence of movements
  - Automatic clustering

(Warnock & Lalmas, 2013)

STUDY III: SELF-REPORT AND BEHAVIOURAL DATA

- **Information Visualization System**
  - McGill Library Catalogue: Engineering Subject Area
    - Version 1: visualization
    - Version 2: visualization + audio

- **Participatory Design Study**

- **Experiment**
  - *n*=24 engineering students
  - Tasks: six information retrieval and hierarchical navigation tasks
  - Data collected: self-report and performance metrics

(Absar, 2012)

FINDINGS

- No difference in performance accuracy or time on task
- Aesthetics and Perceived Usability was higher for the audio-visual system.
- Perceived ease of use was also rated higher for the audio-visual system.
- Open-ended comments offered insights into participants' perceptions and interactions.

STUDY IV: ONLINE NEWS INTERACTIONS

(O'Brien & Lebow, in press)

http://www.cbc.ca/news/

SELF-REPORT, BEHAVIOR AND PHYSIOLOGICAL DATA: MEASURES

- **Pre-task questionnaire**
  - Demographics + news behaviours
- **Interaction with website**
  - Performance**:** Time on task, reading time, browsing time, number of pages visited within site, whether participants clicked on links to recommended content
  - Physiological: heart rate (HR), electrodermal activity (EDA), electrocmytogram (EMG) [subset of participants]
- **Post-session questionnaire**
  - User Engagement Scale (UES) (O'Brien & Toms, 2010)
  - Cognitive Absorption Scale (CAS) (Argawal & Karahanna, 2000)
  - System Usability Scale (SUS) (Brooks, 1997)
- **Think-After Interview**
  - Questions about the items selected for the task
  - Questions about overall experience

SELF-REPORT, BEHAVIOR AND PHYSIOLOGICAL DATA: RESULTS

- **Self-report UES, CAS and SUS**
  - Positive correlations support criterion validity of the measures
  - Designation of “low,” “medium” and “high” scores for each group based on median
  - All questionnaires were positively correlated with aggregate interest in the articles

- **UES and Physiological Data**

	HR	EDA	EMG
UES	-0.38	-0.25	-0.21

SELF-REPORT, BEHAVIOR AND PHYSIOLOGICAL DATA: RESULTS

- **UES and Behavioural Data**

	High M(SD)	Medium M(SD)	Low M(SD)	Kruskal Wallis (x^2)	*p*
Reading time	6:03 (2:34)	6:05 (1:56)	6:56 (3:29)	1.15	0.56
Browsing time	4:03 (2:29)	5:17 (3:49)	7:29 (4:09)	3.98	0.13
Total time	10:07 (3:37)	11:23 (5:10)	14:26 (5:02)	5.09	0.07
# pages visited	9.5 (5.0)	10.3 (3.6)	16.3 (8.4)	3.89	0.14

- Use of Links
  - UES scores were not significantly different between those who clicked on links (M=3.8, SD=0.95) and those who did not (M=4.29, SD=0.52)
  - $U(1)=51.5$, $p=0.15$

THINK-AFTER INTERVIEW

- **Did participants' experiences with online news fit the process model of user engagement** (O'Brien & Toms, 2008)**?**

- **What attributes of user engagement were significant to participants in the online news environment?**
  - Novelty, affect, usability, personal interest and relevance

- **Evidence of two types of engagement** (O'Brien, 2011)
  - Content engagement
  - Interface engagement

OUTLINE

- **Introduction and Scope**

- **Part I - Foundations**
  1. Approaches based on self-report measures
  2. Approaches based on web analytics
  3. Approaches based on physiological measures

- **Part II – Advanced Aspects**
  1. Measuring user engagement in mobile information searching
  2. Networked user engagement
  3. Combining different approaches

- **Conclusions**
- **Bibliography**

CONCLUSIONS

AN EXAMPLE OF PUTTING IT ALL TOGETHER

- Based on experience in working with user-centered products
- Not all measures appropriate to all products

1. articulate the goal(s) of an application/feature

2. identify signals that indicate success

3. build/chose corresponding measures to track

Measures should relate to one or several goals of the application/product
Measures should be used to track progress towards that goal

(Rodden, Hutchinson & Fu, 2010)

GOALS – SOME TIPS

- **What are the goals of the product/features in terms of user experience (user engagement)?**
- **What tasks users need to accomplish?**
- **What is the redesign cycle trying to achieve?**

- **Retention or adoption:**
  - Is it more important to acquire new users or to keep existing ones more engaged?

- **Goal associated with a feature is not the same as goal of the whole product**

- **Measures (to be used or developed) should not be used to solely drive the goals**

SIGNALS – SOME TIPS

- **What is success? What is failure?**
- **What feelings and perceptions correlate with success and failure?**
- **What actions indicate that goals are met?**

- **Data sources**
  - logs, surveys, panel of judges

- **Sensitive and specific signals**
  - need to observe some reaction when user experience is better or worse
  - failure often easier to identify than success
    - undo event, abandonment, frustration

MEASURES – SOME TIPS

- **Raw counts need to normalised**
- **Ratios, percentages, and average per users often more useful**

- **Accuracy of metrics**
  - bots, all important actions recovered

- **Keep comparing measures with “conventional” ones (e.g. comScore matters)**

(Rodden, Hutchinson & Fu, 2010)

OPEN RESEARCH QUESTIONS

… IN NO PARTICULAR ORDER

- A great deal of emphasis on users and systems, but less evidence about the role of task, device, and context on user engagement.
- We tend to focus on characteristics of users *in the moment* of interaction. But are their individual differences that may predict the level of engagement that can be achieved?
- Psychophysiological measurement may not be sensitive enough for measuring "general" or "average" engagement (e.g. News or Mail sites) … although it will likely bring great insights.
- How we to "use" physiological measures – interpretation of the data generated – is an important area for exploration.
- For any measurement that we "think" may be important (e.g. cursor vs. relevance), we need to made explicit connections to engagement

- Be careful of the WEIRD syndrome ((Western, Educated, Industrialized, Rich, and Democratic)

CONCLUSIONS

- We covered a range of self-report, performance and physiological metrics.
- We focused on different characteristics of measures, including intra- vs. inter-session; subjective vs. objective; process- vs. product-based, small- vs. large-scale; and lab vs. field.

Take-Aways

- No one measure is perfect or complete.
- All studies have different constraints.
- More details on methods used in published literature will enhance communication around UE measures, which will advance study of UE.
- Need to ensure methods are applied *consistently* with attention to reliability.
- More emphasis should be placed on using mixed methods to improve the validity of the measures.

ACKNOWLEDGEMENTS

- Dr. Lalmas work in collaboration with Ioannis Arapakis, Ricardo Baeza-Yates, Berkant Cambazoglu, Georges Dupret, Janette Lehmann and others at Yahoo! Labs.

- Dr. O'Brien's work is supported by the Social Science and Humanities Research Council (SSHRC) of Canada and the Networks of Centres of Excellence Graphics, Animation and New Media (NCE GRAND) Project (http://www.grand-nce.ca/).