

**field studies, emotion, and affect**

dr. carman neustaedter


**field studies**



naturalistic studies  
in context, *in situ*. Culture, history and folkways

- **ethnographic:** early user research – adoption of technologies
- **field evaluation:** summative evaluation of a product in real situations, use

**field studies**



**pros:**  
natural, real behavior  
good if you can't replicate the setting

**cons:**  
easy to intrude or intervene  
hard to replicate sometimes

**usability (nielsen's definition)**

**learnability:** easy to learn so a user can rapidly start to use it

**efficiency:** once the user has learned the system, a high degree of productivity is possible

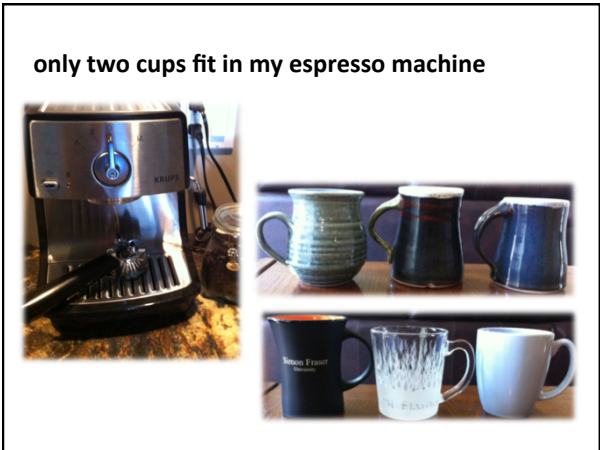
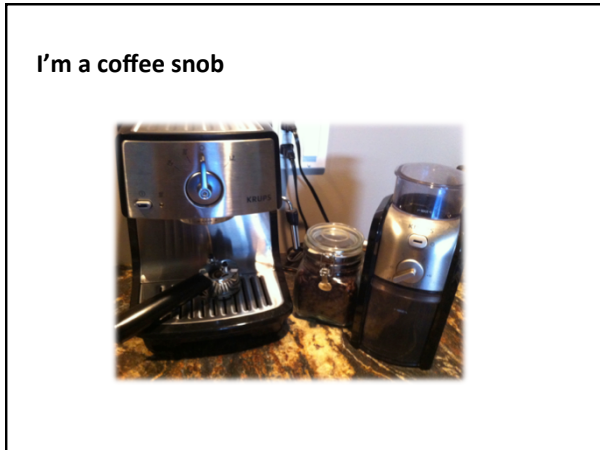
**memorability:** the user should be able to return to the system and not have to learn again

**errors:** users should make few errors and recover easily

**satisfaction:** the system should be pleasant to use

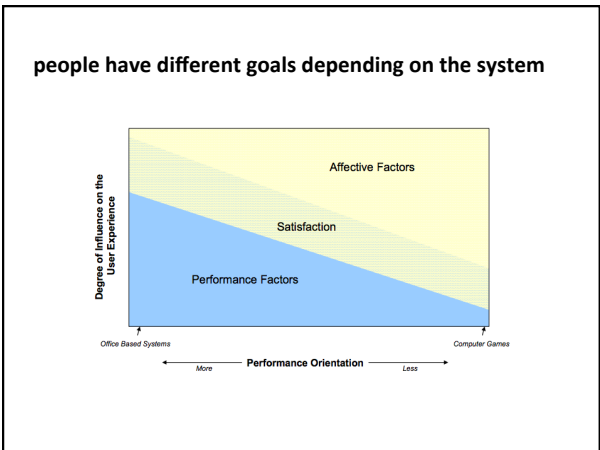
**is there something more?**





**the problem**

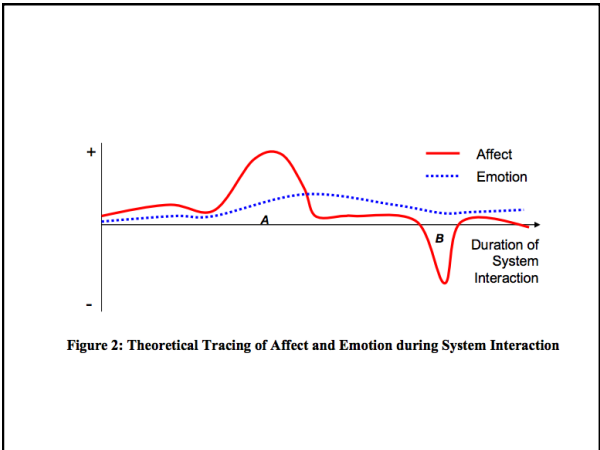
Traditional design evaluation focus on **performance**, but neglects **people's emotional responses**



**definitions**

**emotion:** overall feeling, summative, reflective influenced by **affect**.

**affect:** short term, immediate, non-reflective influences on **emotion**. **Can be volatile**



**theories of emotion**

agreement on three components:

1. **cognitive/experiential** (subjective feeling)
2. **physiological response** (body's internal response)
3. **behavioral response** (body's external response)

**measuring it: cognitive/experiential**

- use interview or questionnaire **after the fact**
- this will only get **emotion**; lack of feelings 'in the moment' (affect)
- self-reporting can be subjective

**intrinsic motivation inventory (IMI) questionnaire**

- measures enjoyment
- you will use it in your assignment
- compare it to cued debrief recall

**Intrinsic Motivation Inventory (IMI)/ Enjoyment Questionnaire**

Evaluator ID: 00-000  
 Game Platform: Wii PS2/EyeToy (circle one)  
 Game: \_\_\_\_\_  
 Participant Number: \_\_\_\_\_  
 Date and time of session: \_\_\_\_\_

Player to answer:

For each of the following statements, please indicate how true it is for you, using the following scale:

1	2	3	4	5	6	7
not at all true			neutral true			very true

1. While I was working on the task I was thinking about how much I enjoyed it.
2. I did not feel at all nervous about doing the task.
3. I felt that it was my choice to do the task.
4. I think I am pretty good at this task.
5. I found the task very interesting.
6. I felt tense while doing the task.
7. I think I did pretty well at this activity, compared to other students.
8. Doing the task was fun.
9. I felt relaxed while doing the task.
10. I enjoyed doing the task very much.
11. I didn't really have a choice about doing the task.
12. I am satisfied with my performance at this task.
13. I was anxious while doing the task.

**scoring**

**Scoring Information for IMI/Enjoyment Questionnaire**

**Scoring information.** Begin by reverse scoring items # 2, 9, 11, 14, 19, 21. In other words, subtract the item response from 8, and use the result as the item score for that item. This way, a higher score will indicate more of the concept described in the subscale name. Thus, a higher score on pressure/tension means the person felt more pressured and tense; a higher score on perceived competence means the person felt more competent; and so on. Then calculate subscale scores by averaging the items scores for the items on each subscale. They are as follows. The (R) after an item number is just a reminder that the item score is the reverse of the participant's response on that item.

Interest/enjoyment: 1, 5, 8, 10, 14(R), 17, 20  
 Perceived competence: 4, 7, 12, 16, 22  
 Perceived choice: 3, 11(R), 15, 19(R), 21(R)  
 Pressure/tension: 2(R), 6, 9(R), 13, 18

Determine the average subscale score and standard deviation for each participant.


**Results**

Participant 1	SubScale Name	Average Rating*	Std Deviation**
	Interest/enjoyment		
	Perceived competence		
	Perceived choice		
	Pressure/tension		

\* Average of scores for all items/questions related to each subscale name for this participant only.  
 \*\* Standard deviation for all items/questions related to each subscale name (use excel)

**measuring it: physiological**

measure physiological changes  
e.g.,  
galvanic skin response (sweat)  
heart rate  
breathing rate



could measure affect

but you have to interpret the sensed data  
e.g., does increased heart and breathing rate mean fear or excitement?

**measuring it: behavioral**

measure behavioral changes  
e.g.,  
facial expressions  
changes in tone of voice  
gestural activity



could measure affect

but involves interpretation of data  
e.g., is that facial expression showing extreme happiness or sadness?

**measuring emotion & affect**

**emotion** is easy to ask at the end of a study  
**affect** is harder to gather during the study – transient.

affect could be done through think aloud, but this interrupts and may not be accurate

**cued recall debrief (aka “aided recall method”)**


interview technique based on **review** or **situated recall**

1. **record** a first person view
2. **play it back** to be re-immersed in the activity
3. **interview** during playback and record *third* person view; ask them to describe their feelings
4. **analyze** the final video: analyze comments related to affective experience

**first person view**



**third person view of the interview**






**why count the neutral?**

<b>person one</b> 20 positive 5 negative	<b>person two</b> 20 positive 5 negative
------------------------------------------------	------------------------------------------------

similar experience?


<b>person one</b> 20 positive 5 negative 75 neutral = 20% positive <i>neutral provides additional context</i>	<b>person two</b> 20 positive 5 negative 5 neutral = 66% positive
------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------

**sonic the hedgehog example** 

**Note:** this quote is of game play, you need to analyze the debrief audioed debrief recall

“This is hard. Hmmm. How do you get the ... Ok, I’ve got it. Cool, got it. Oh, no I missed again. Shoot, I’m gone. This is great. Ok, I’m trying again. Oh, not that pesky bug again. Got ya. I’m having fun. No just let me finish.”


how many positive and negative affective comments?  
how many neutral?

**sonic the hedgehog example** 

“This is hard. Hmmm. How do you get the ... Ok, I’ve got it. **Cool, got it.** Oh, no I missed again. Shoot, I’m gone. **This is great.** Ok, **I’m trying again.** Oh, not that pesky bug again. **Got ya. I’m having fun.** No just let me finish.”

positive (5)


- “Cool, got it” (feeling good)
- “This is great” (feeling good)
- “I’m trying again” (not giving up)
- “Got ya” (success/winning)
- “I’m having fun” (fun)

**sonic the hedgehog example** 

“**This is hard.** Hmmm. How do you get the ... Ok, I’ve got it. Cool, got it. **Oh, no** I missed again. Shoot, I’m gone. This is great. Ok, I’m trying again. **Oh, not that pesky bug again.** Got ya. I’m having fun. **No just let me finish.**”

negative (5)

- “This is hard” (hard/frustrated)
- “Oh no” (feeling bad)
- “I’m gone” (feeling bad/failure)
- “Oh, not that pesky bug again” (frustrated)
- “No just let me finish” (not ready to quit/frustrated)

**sonic the hedgehog example** 

“This is hard. Hmmm. How do you get the ... **Ok, I’ve got it.** Cool, got it. Oh, no I missed again. Shoot, I’m gone. This is great. Ok, I’m trying again. Oh, not that pesky bug again. Got ya. I’m having fun. No just let me finish.”

neutral (1)

- “Ok, I’ve got it.”

**coding is subjective**

how to ensure coding is **reliable/consistent?**

**reliable:** repeatable by others  
**consistent** between coders/evaluators  
**consistent** across all participants

calculate **inter-rater reliability**

### inter-rater reliability

- at least two coders/evaluators
- use same definition of types of codes
- code separately
- compare results

### “perfect” example

<b>evaluator A</b>	<b>evaluator B</b>
positive 8	positive 8
negative 4	negative 4
neutral 6	neutral 6

notice the slope of this line is 1; the correlation is 1

### not “perfect” example

<b>evaluator A</b>	<b>evaluator B</b>
positive 4	positive 8
negative 2	negative 6
neutral 3	neutral 6

### correlation

the relationship between two variables  
an indication of the strength of the relationship

**correlation coefficient:** from -1 to 1

- 1 = perfect relationship
- 0 = no relationship
- 1 = opposite relationship (e.g., as one goes up, the other goes down)

### inter-rater reliability

- use correlation coefficient
- compare ratings/coding of two evaluators

if  $r > 0.9$ , then they are reliably the same (81% of rater series 2 is known from rater series 1). Depending on sample size, this will likely be stable or hold above chance > 19 of 20 trials.

### correlation example in Excel

Inter-rater Reliability Using Correlation

Combine number of each type of comments for all participants. So, in this example, there are only 15 comments made by all participants.

Comments Type	EXAMPLE 1		EXAMPLE 2	
	Rater A	Rater B	Rater A	Rater B
Positive	10	6	10	8
Neutral	2	6	2	6
Negative	3	3	3	5
Total Comments	15	15	15	19

R = 0.8 Not good inter-rater reliability      R = 0.8 Good inter-rater reliability

Correlation R = 1.0 is perfect correlation between two sets of data.  
R = 0.9 is acceptable. Chances, go back and review again together to reach a better consensus.  
I ran correlation (CORREL) on the grey data for each case (the good and not good case).

**wrapping up**