PREFERENCE TESTING

IN USER EXPERIENCE RESEARCH: TO DO OR NOT TO DO?

MOLLY MALSAM, SENIOR USER EXPERIENCE RESEARCHER

ABOUT ME

- Went to graduate school intending to work in politics…
- Technical writer for 12 years
- User Experience field for 16 years
- Dedicated to User Experience Research for almost 4 years at Charles Schwab
- Worked in a lot of different industries and companies of many sizes, but primarily focused on digital experiences

AGENDA

- The definition of preference testing
- How valuable is preference testing as part of UX research
- Case studies
- Preference testing guidelines

WHAT IS PREFERENCE TESTING?

WHAT IS PREFERENCE TESTING?

The way I most often hear people define a preference test when speaking of usability research…

Showing several design options to participants and asking them to choose their favorite

COMMON QUESTIONS

A Usability Hub post lists these commonly used questions in preference tests:

- Which design does a user prefer?
- Which design looks the most trustworthy?
- Which design looks the easiest to use?

Other online instructional guides focus on the idea of asking users what *seems* more usable rather than actually testing its usability

Source: https://tinyurl.com/23aduywr

OFTEN HAS A VISUAL / AESTHETIC FOCUS

These types of tests often focus on the visual / aesthetic appeal of a design now that things like fonts, colors, and images are increasingly expected to express complex brand traits like friendliness or trustworthiness

NN/g details how to do preference testing on these kinds of aesthetic impressions

Source: https://www.nngroup.com/articles/testing-visual-design/

OFTEN HAS A VISUAL / AESTHETIC FOCUS

Aesthetically-focused testing is **not** the focus of this talk

While some may debate whether that type of research is truly usability testing, I want to focus on traditional usability testing…

Which is the evaluation and measurement of a product or service by having participants complete typical tasks with it…

And using that format to ask questions about what users *prefer or choose* when showing more than one design option

IN OTHER WORDS…

If you are a User Experience Researcher and you have been asked to, or want to, evaluate user preference in some way as part of your research

Also, the focus here is on digital products / interfaces

HOW VALUABLE IS PREFERENCE TESTING AS A PART OF USABILITY RESEARCH?

POLL QUESTION

When sharing two or more designs, do you think it's a good idea to ask usability test participants which design they like better?

- Yes
- No
- I don’t know

WHERE DID PREFERENCE TESTING COME FROM?

- In the early 70s, scientists used preference tests as a means of answering questions about animal welfare
- Wikipedia also focuses on its earliest usage in animal behavior and motivation
- In consumer sciences and marketing, preference tests such as taste tests have been around for decades
- It's unclear when it started being used in the UX field

Source: https://en.wikipedia.org/wiki/Preference_test

WHY DO WE PREFERENCE TEST?

Asking people's preference has face validity – we assume that what people prefer will **also align with other positive things** we want to impact, like:

- an easier, more efficient, and/or more satisfying user experience
- "converting" in some way / improving sales
- forming a favorable opinion of the brand
- enhancing trust

TRUE OR FALSE?

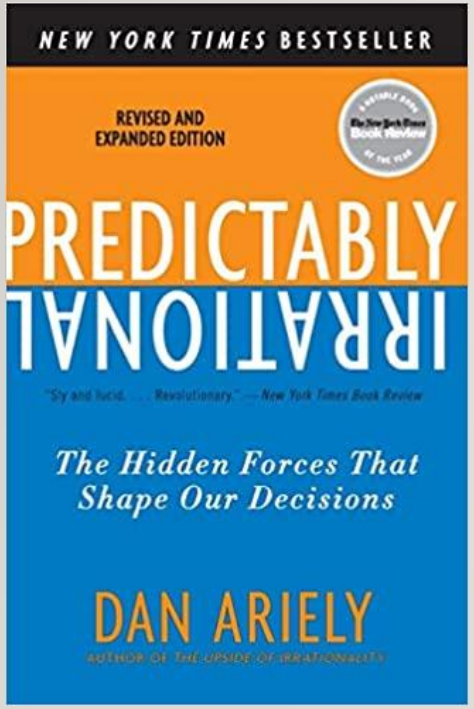
But how do we know that is true?

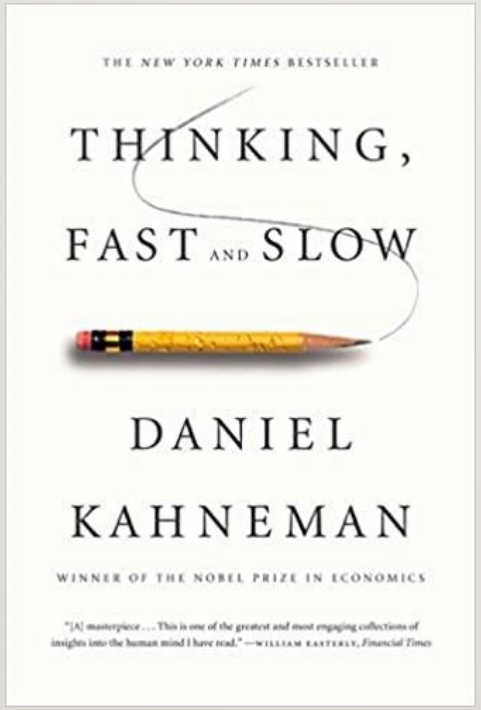
As it turns out, research in the psychology field shows we don't have a clear understanding of ***why*** people choose what they choose, much less what that choice may or may not predict

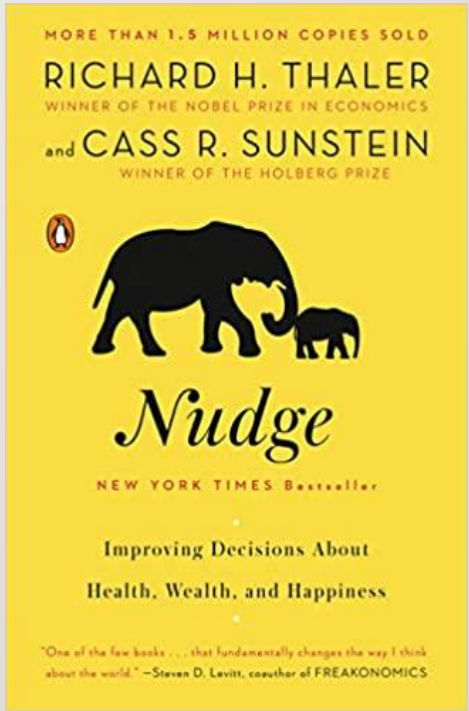
NOT FOUND TO BE HIGHLY VALUABLE

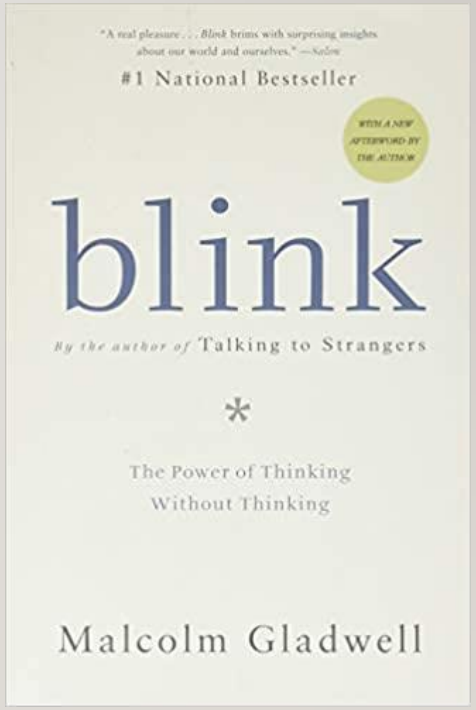
I'm going to argue that the existing evidence indicates that preference questions *do not* provide value to the digital product UX evaluation process, and may result in less usable experiences

BESTSELLERS ON DECISION-MAKING









FRAMING EFFECTS

A cognitive bias where people's responses are influenced by the way a piece of information is presented to them

This is one of several in a general class of judgement and decision-making fallacies identified by researchers

Source: https://en.wikipedia.org/wiki/Framing_effect_(psychology)

FRAMING EFFECTS: POSITIVE VS. NEGATIVE

You have a 90% chance of surviving the operation

You have a 10% chance of dying during the operation

Source: Moxey, O'Connell, McGettigan, and Henry, "Describing Treatment Effects to Patients", 2003. *J. Gen Intern Med.*

FRAMING EFFECTS: POSITIVE VS. NEGATIVE

- You have a 90% chance of surviving the operation

You have a 10% chance of dying during the operation

Source: Moxey, O'Connell, McGettigan, and Henry, "Describing Treatment Effects to Patients", 2003. *J. Gen Intern Med.*

FRAMING EFFECTS: POSITIVE VS. NEGATIVE

Source: https://tinyurl.com/atuje6ee

FRAMING EFFECTS: LEANER IS BETTER

Source: https://tinyurl.com/atuje6ee

FRAMING EFFECTS: VALUE

For

$290

BUY NOW

Now only:

~~$400~~ **$290**

BUY NOW

Source: https://livesession.io/blog/cognitive-bias-ux/

FRAMING EFFECTS: GETTING A DEAL

For

$290

BUY NOW

Now only:

~~$400~~ **$290**

BUY NOW

Source: https://livesession.io/blog/cognitive-bias-ux/

CHOICE BLINDNESS

An observable psychological phenomenon in which even when someone doesn't get what they want, there's a strong chance they won't even notice, and may even defend a choice they think that they made

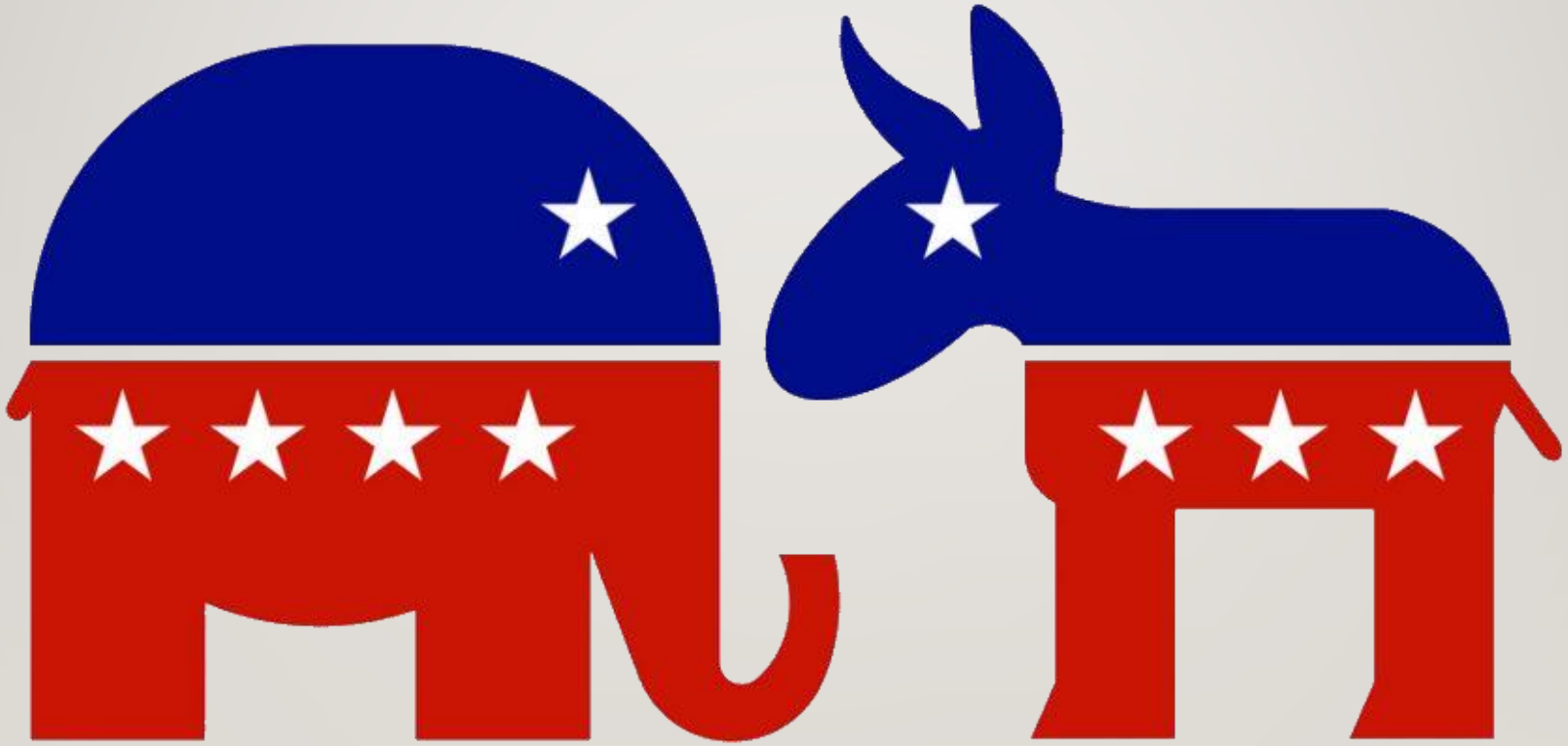
CHOICE BLINDNESS: FACIAL PREFERENCE

Source: https://portal.research.lu.se/portal/files/4436175/547323.pdf

CHOICE BLINDNESS: JAM PREFERENCE

Source: Hall, Johannson, et al ,"Magic at the marketplace", 2010. *Cognition,* 117.

CHOICE BLINDNESS: POLITICAL ISSUES



SUBSTITUTION EFFECT

If we can’t come up with a choice quickly, we find a related question that is easier to answer and answer it instead

SUBSTITUTION EXAMPLES

Target Question	Substitute Question
How much would you contribute to save an endangered species?	How much emotion do I feel when I think of dying dolphins?
How happy are you with your life these days?	What is my mood right now?
How popular will the president be in 6 months?	How popular is the president right now?
How should financial advisors who prey on the elderly be punished?	How much anger do I feel when I think of financial predators?

Source: Kahneman, "Thinking, fast and slow", 2011.

SUBSTITUTION IN DESIGN CHOICES?

Target Question	Substitute Question
Which design do you like better?	Could it be…. • Which one has my favorite color? • Which one has an image I react to more favorably? • Which one looks shorter? (but doesn't give enough information to complete the task successfully?)

AESTHETIC-USABILITY EFFECT

People are more tolerant of minor usability issues when they find an interface visually appealing

This means that a user could choose a design that is less effective in terms of task completion based on aesthetics

Source: https://tinyurl.com/wwaak335

AESTHETIC-USABILITY EFFECT: EXAMPLE

In a test done by the NN/g, a user experienced several usability issues, from minor annoyances to serious navigation problems

But in a post-task questionnaire, she rated her experience in terms of ease of use very highly, commenting that:

> "It's the colors they used," she said. "Looks like the ocean, it's calm. Very good photographs."

Source: https://www.nngroup.com/articles/aesthetic-usability-effect/

THE DECOY EFFECT

A phenomenon in which people change their preference between two options when presented with a third option

The third option is considered much less attractive than the other two options, which is called "asymmetrically dominated"

Source: https://doi.org/10.1086/208899

DECOY EFFECT: EXAMPLE

Source: http://humanhow.com/the-decoy-effect-complete-guide/

DECOY EFFECT: EXAMPLE

Source: http://humanhow.com/the-decoy-effect-complete-guide/

CONCLUSION

Given all these ways that choices can be easily influenced, manipulated, or just not based on anything particularly **rational,** why do we want to measure them in our usability tests?

UX researchers support user efficiency, effectiveness, and satisfaction (usability defined)

We already have valid and reliable ways of measuring usability that don't have anything to do with people's choices

CASE STUDIES

PREFERENCE TEST EXAMPLES

Both of these studies were designed to:

- Be as close as possible to the sample size needed to detect at least a 10% difference between options, if one exists
- Have participants evaluate the options both independently, based in usability principles, and comparatively

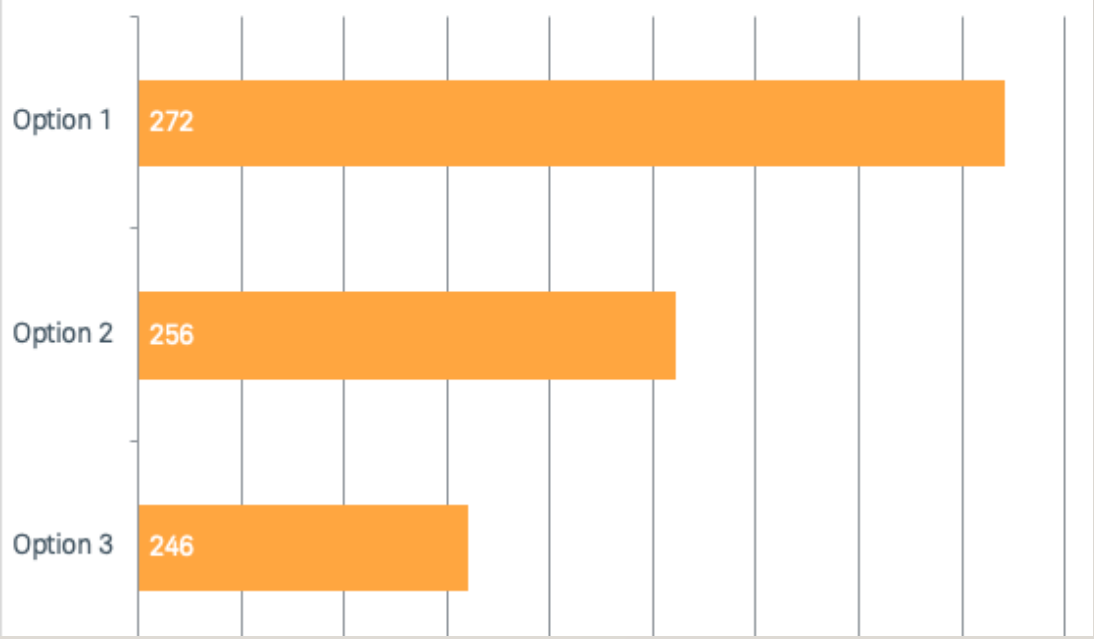
PREFERENCE TEST EXAMPLE 1: BACKGROUND

Three design options were presented for guiding users to products that matched their needs

Within-subjects study with 129 participants

PREFERENCE TEST EXAMPLE 1: COMPARATIVE RESULTS

When asked to choose the one they preferred, Option 1 was the clear winner



PREFERENCE TEST EXAMPLE 1: INDEPENDENT RESULTS

However, when I measured ease of use, understanding, and satisfaction after each option:

Option 2 was rated, on average, more highly than Option 1, even though they chose option 1

PREFERENCE TEST EXAMPLE 1: DISCUSSION

Why the difference?

- Looking more closely at the data set, those with less investing experience both preferred & rated Option 2 higher, just not as strongly whereas those with more investing experience showed the disparity and more strongly preferred Option 1
- The reasons given for that set who preferred Option 1 but rated Option 2 more highly were largely visual (liked having icons, overall layout, a particular interactive behavior, "cooler")

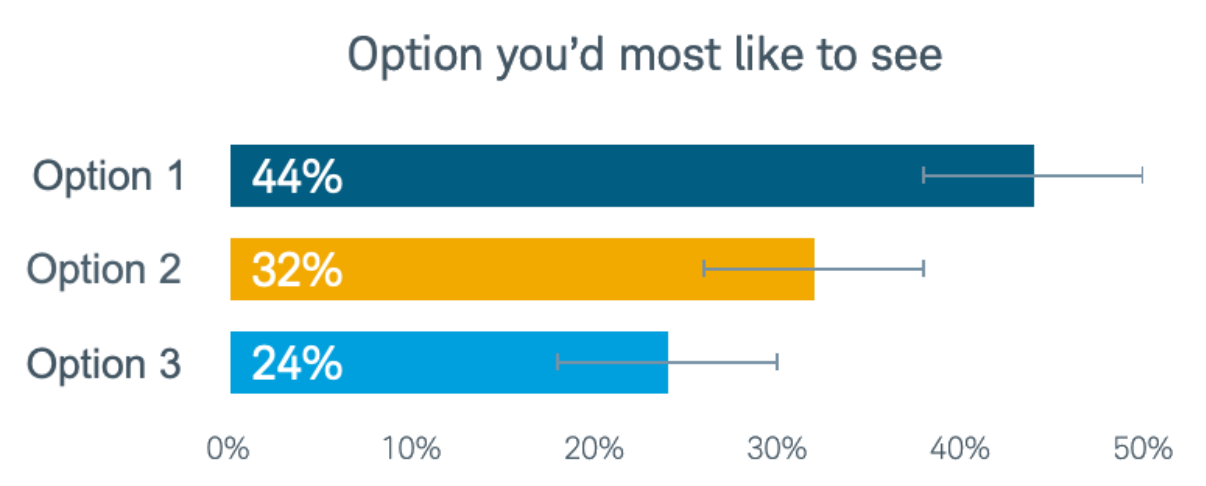
PREFERENCE TEST EXAMPLE 2: BACKGROUND

Three design options were presented for displaying projected income from an investment over a given timeframe, as a challenge to the hegemony of line charts

Within-subjects study with 160 participants

PREFERENCE TEST EXAMPLE 2: COMPARATIVE RESULTS

When asked to choose the one they preferred, Option 1 was the clear winner (due to data type we were able to include confidence intervals here)



PREFERENCE TEST EXAMPLE 2: INDEPENDENT RESULTS

However, when asked after each option to independently evaluate understanding, confidence, and satisfaction:

- Options 2 and 3 were rated significantly easier to understand
- Option 3 inspired significantly higher levels of confidence than Option 1, with Option 2 higher but not significantly
- Option 3 had non-significant higher levels of satisfaction

PREFERENCE TEST EXAMPLE 2: DISCUSSION

Why the difference?

- Segmenting out the data set on key variables showed no difference to the pattern
- The reasons given for that set were largely based on their preference for data like this in a chart format

CONCLUSION

This is only two studies so it's not solid proof but..

In both cases, one design was rated higher on independent variables related to usability than what users preferred, providing support that preference doesn't correlate with usability

Maybe future studies will find positive correlations with other outcomes that would be considered desirable, so that is an area of research opportunity

POLL QUESTION

When sharing two or more designs, do you think it's a good idea to ask usability test participants which design they like better?

- Yes
- No
- I don’t know

IF YOU'RE STILL GOING TO (OR HAVE TO) DO IT…

Here are some guidelines

GUIDELINE #1: ASK FOR CHOICE EXPLANATION

After asking users what they like or prefer, ask them to explain

Carefully analyze these results to separate responses that don't relate to anything that truly impact usability and that likely come from one or more of the decision-making effects we discussed

Warning: As mentioned, people don't always **know** why they make a choice, so this is a "better than not doing it at all" recommendation for those who must preference test

GUIDELINE # 2: MEASURE OTHER DEPENDENT VARIABLES

Don't *just* ask which one users like or prefer

Make sure to do tasks and measure other dependent variables you want to impact with your design options:

- Efficiency – time on task
- Effectiveness – task success rate, error rate
- Overall usability and ease of use – (ex. SUS, SEQ)
- Satisfaction (ex. CSAT)

GUIDELINE #3: MEASURE OPTIONS INDEPENDENTLY

Rather than say, measuring ease of use by asking users to choose the one that was easier…

Ask SEQ after each option and then compare the results to look for significant differences (typically requires unmoderated higher-volume testing)

GUIDELINE #4: INDEPENDENT MEASURES OVERRULE

If comparative results either:

- Show some significant difference between the options that the independent measures do not
- Show a significant difference that is not the same as a significant difference found in the independent measures

The independent measures should be the deciding factor

SUMMARY

- The definition of preference testing
- How valuable is preference testing
- Case studies
- Preference testing guidelines

QUESTIONS


Can you repeat the part of the stuff
where you said all about the things?