

# **Building User Interfaces**

# **Designing Agents**

## **Cole Nelson**

# What will we learn today?

- How can we understand computers as social actors?
- How can we design character speech?
- What is personality in artificial agents?

# Computers as Social Actors

## Computers as Social Actors (CASA)<sup>1</sup>

**Definition:** A paradigm that states that humans *mindlessly* apply the same social heuristics used for human interactions to computers and treat them as social agents.

**Mindlessness** is an inactive state of mind that is characterized by reliance on distinctions drawn in the past.

<sup>1</sup> Nass & Moon, 2000, Machines and Mindlessness: Social Responses to Computers

# CASA extends to many social-psychological concepts.<sup>1</sup>

Concept	Evidence
Gender	People mindlessly gender-stereotype computers.
Ethnicity	People favor computers with similar ethnicity cues.
Group membership	People are more collaborative with computers that are in their team.
Politeness	People show politeness toward computers that socially engage with them.
Reciprocity	People help a computer that was helpful to them.
Personality	People are attracted to computers with similar personality.

<sup>1</sup> Nass & Moon, 2000, Machines and Mindlessness: Social Responses to Computers

## **Similarity-Attraction Theory<sup>2</sup>**

**Definition:** A social-psychological theory that posits that people like and are attracted to agents that are similar, rather than dissimilar, to themselves.

**Likeness begets liking**

<sup>2</sup> Byrne et al, 1967, Attraction and similarity of personality characteristics.

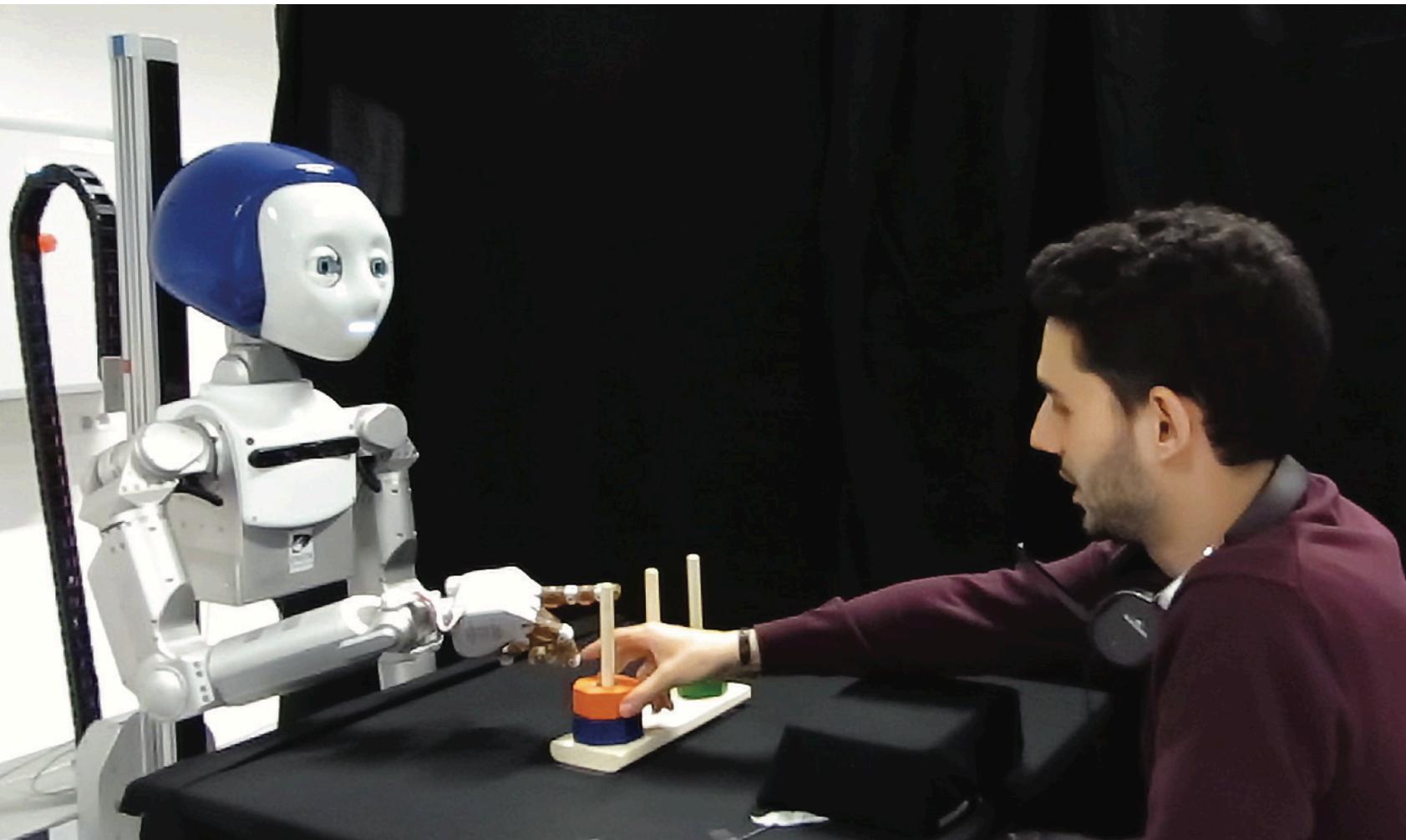
## Similarity-Attraction in Character Design

- Matching the user in explicit characteristics
  - **Visible:** age, gender, clothing
  - **Behavioral:** language, accent
- Matching the user in implicit characteristics
  - **Personality:** extroversion, agreeableness
  - **Interaction style:** formal, informal
- Matching user preferences

## Example of Similarity-Attraction<sup>3</sup>

A robot coach matched the personality of its user purely by increasing/decreasing eye contact.

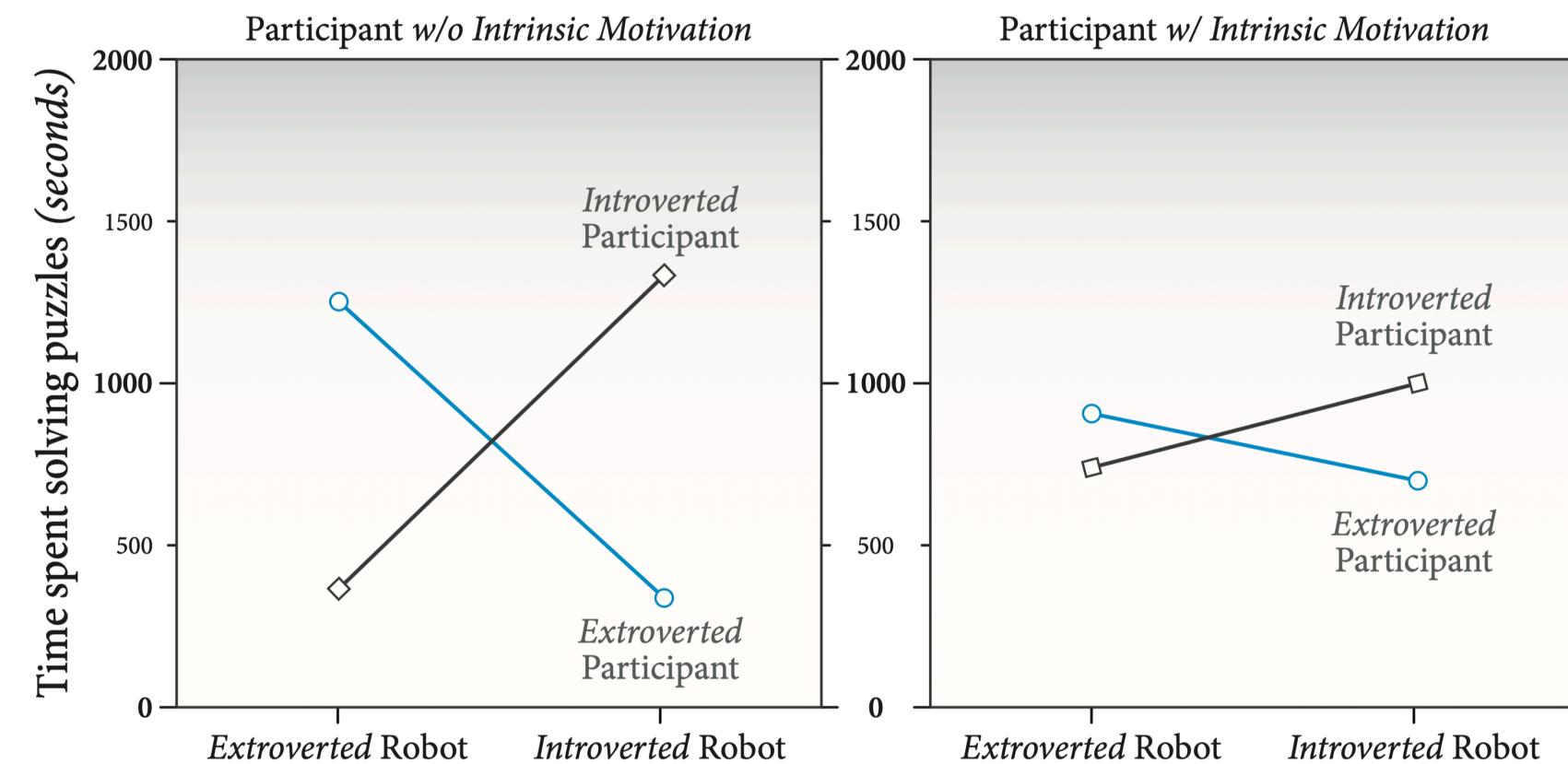
- Extroverts build more eye contact; introverts build less eye contact.



<sup>3</sup> Andrist et al., 2015, Look Like Me...

Results support similarity-attraction theory: people were motivated by a robot with a similar personality.<sup>3</sup>

- Although intrinsic motivation trumps any motivation that an agent can provide.



<sup>3</sup> Andrist et al., 2015, Look Like Me...

## Consistency-Attraction Theory<sup>4</sup>

**Definition:** People prefer to interact with agents that behave consistently, rather than inconsistently. Consistency reduces cognitive load, makes it easier to predict what will happen next.

**Internal consistency:** The behaviors, appearance, function, and so on are consistent with each other.

**External consistency:** The design is consistent with the expectations and preferences of the user.

<sup>4</sup> Nass & Lee, 2001, Does computer-synthesized speech manifest personality?

## Consistency-Attraction Example<sup>5</sup>

To create *believable* characters:

- following a dramatic structure with rich backstory and evolving story line
- utilizing verbal and nonverbal social behaviors
- expressing culture



<sup>5</sup> Simmons, 2011, Believable robot characters

# Designing Character Speech

## Politeness Theory<sup>6</sup>

**Definition:** *Politeness theory* posits that individuals utilize communication strategies that express concern for others and minimize threats to self-esteem.

<sup>6</sup> Brown & Levinson, 1987, Politeness: Some universals in language usage.

**Positive face** is the need for self-image to be accepted, appreciated, and approved of by others.

**Negative face** is the need to be independent, to have freedom of action, and not to be imposed on by others.

**Positive politeness:** avoiding offense through friendliness.

**Negative politeness:** avoiding offense through deference.

**Face saving:** showing deference, prioritizing the other's time or concerns, and including an apology for imposition, when oriented toward negative face; and showing solidarity and sharing of goals, when oriented toward positive face.

**Face threatening:** opposing to the wants/desires of the other.

# Strategies for Positive Politeness<sup>6</sup>

<sup>6</sup> Brown & Levinson, 1987, Politeness: Some universals in language usage.

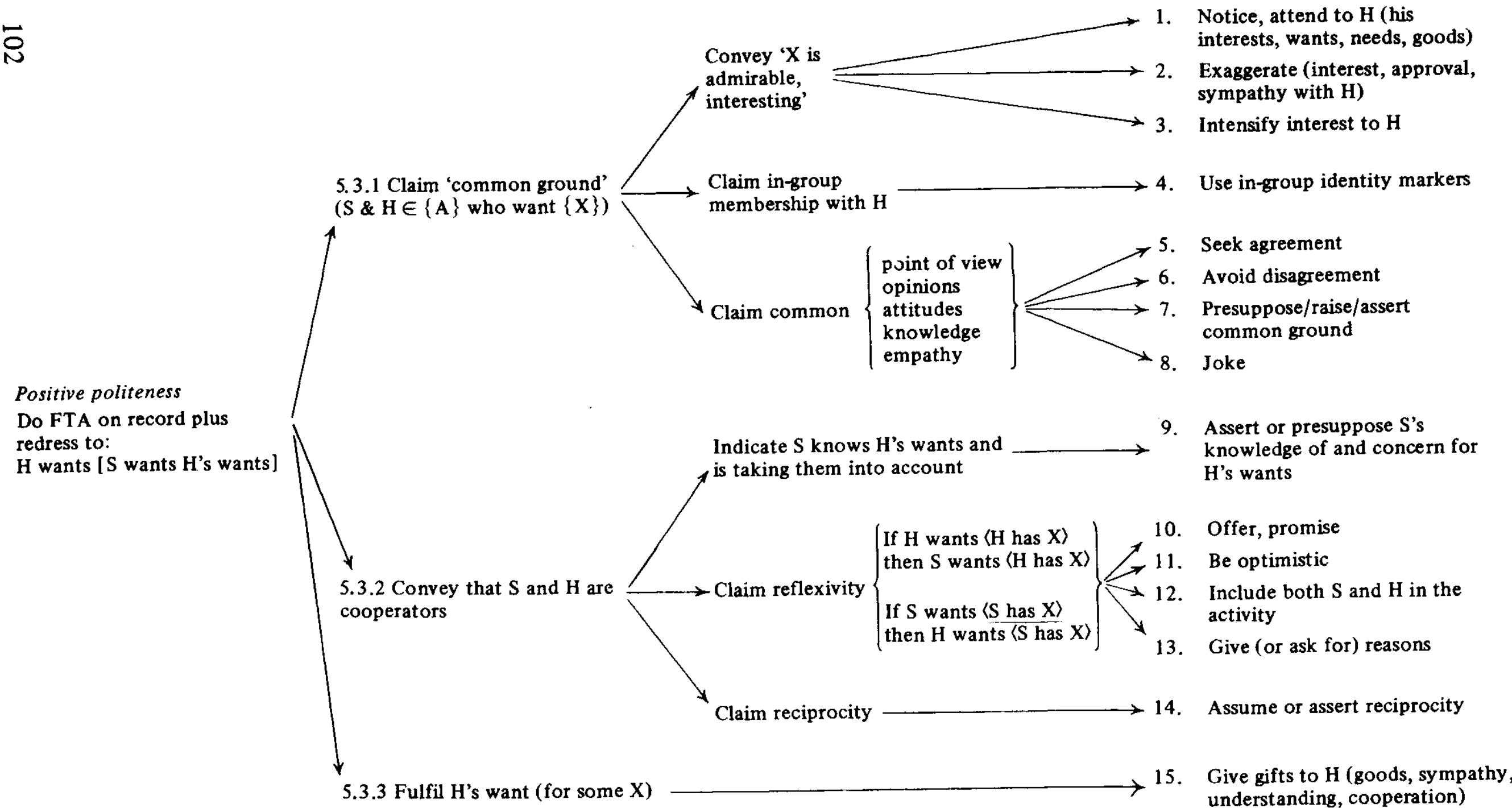


Fig. 3. Chart of strategies: Positive politeness

## Some Examples<sup>6</sup>

**Strategy 1:** Notice, attend to their interests, wants, needs, goods

**Goodness, you cut your hair!** By the way, I came to borrow some flour.

**Strategy 4:** Use in-group identity markers

Help me with this bag here, will you **luv/son/pal?**

**Strategy 5:** Seek agreement through repetition

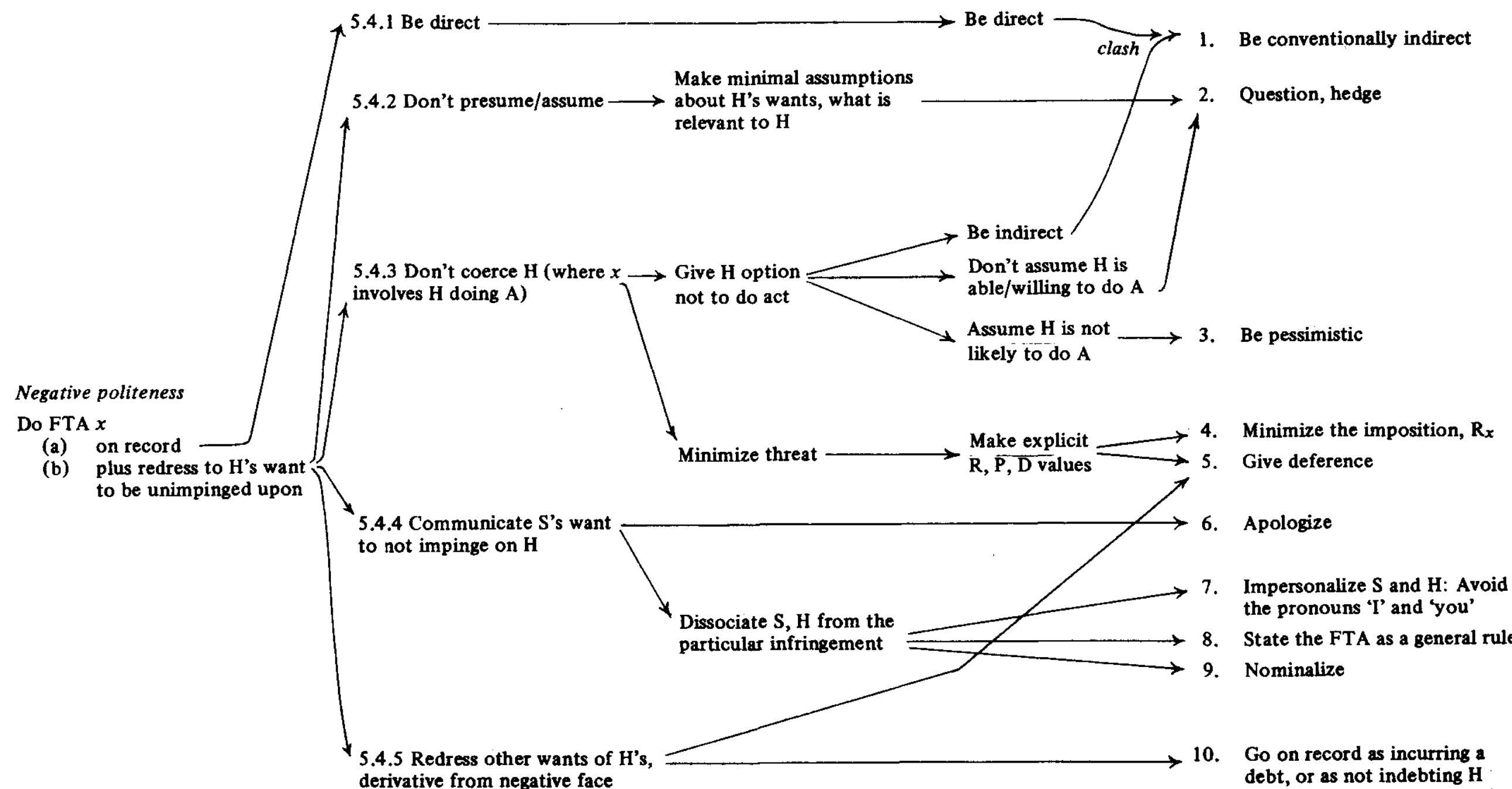
**A:** I had a flat tyre on the way home.

**B:** Oh God, a flat tyre!

<sup>6</sup> Brown & Levinson, 1987, Politeness: Some universals in language usage.

# Strategies for Negative Politeness<sup>6</sup>

<sup>6</sup> Brown & Levinson, 1987, Politeness: Some universals in language usage.



## Some Examples<sup>6</sup>

Strategy 1: Be conventionally indirect

**Are you able to post this letter for me?**

Strategy 2: Question, hedge

I'm **pretty** sure I've read that book before.

A swing is **sort of** a toy.

Strategy 6: Apologize

I hope this isn't going to bother you too much...

<sup>6</sup> Brown & Levinson, 1987, Politeness: Some universals in language usage.

## Hedging<sup>6</sup>

---

Less hedging

---

### Example

---

Lend me your car.

---

May I borrow your car please?

---

I'd like to borrow your car, if you wouldn't mind.

---

Would you have any objections to my borrowing your car for a while?

---

Could you possibly by any chance lend me your car for just a few minutes?

---

More hedging

---

There wouldn't I suppose be any chance of your being able to lend me your car for just a few minutes, would there?

<sup>6</sup> Brown & Levinson, 1987, Politeness: Some universals in language usage.

# Which Strategies Are Most Effective?<sup>7</sup>

Danescu-Niculescu-Mizil and colleagues modeled the relationship between speech strategies and politeness.

Strategy	Politeness	In top quartile	Example
1. Gratitude	0.87***	78%***	I really appreciate that you've done them.
2. Deference	0.78***	70%***	Nice work so far on your rewrite.
3. Greeting	0.43***	45%***	Hey, I just tried to ...
4. Positive lexicon	0.12***	32%***	Wow! / This is a great way to deal...
5. Negative lexicon	-0.13***	22%**	If you're going to accuse me ...
6. Apologizing	0.36***	53%***	Sorry to bother you ...
7. Please	0.49***	57%***	Could you please say more...
8. Please start	-0.30*	22%	Please do not remove warnings ...
9. Indirect (btw)	0.63***	58%**	By the way, where did you find ...
10. Direct question	-0.27***	15%***	What is your native language?
11. Direct start	-0.43***	9%***	So can you retrieve it or not?
12. Counterfactual modal	0.47***	52%***	Could/Would you ...
13. Indicative modal	0.09	27%	Can/Will you ...
14. 1st person start	0.12***	29%**	I have just put the article ...
15. 1st person pl.	0.08*	27%	Could we find a less complex name ...
16. 1st person	0.08***	28%***	It is my view that ...
17. 2nd person	0.05***	30%***	But what's the good source you have in mind?
18. 2nd person start	-0.30***	17%**	You've reverted yourself ...
19. Hedges	0.14***	28%	I suggest we start with ...
20. Factuality	-0.38***	13%***	In fact you did link, ...

<sup>7</sup> Danescu-Niculescu-Mizil et al., 2013, A computational approach to politeness with application to social factors

## Guidelines for Character Design<sup>7</sup>

Most effective *politeness* strategies (do these):

Strategy	Examples
Gratitude	I really appreciate that you've done them
Deference	Nice work so far on your rewrite.
Indirect (btw)	By the way, where did you find...
Please (not start)	Could you please say more...
Apologizing	Sorry to bother you...
Counterfactual modal	Could/Would you...
Greeting	Hey, I just tried to ...

<sup>7</sup> Danescu-Niculescu-Mizil et al., 2013, A computational approach to politeness with application to social factors

# Most effective *rudeness* strategies (don't do these):<sup>7</sup>

Strategy	Examples
Direct start	<b>So can you retrieve it or not?</b>
Factuality	<b>In fact you did link...</b>
2nd person start	<b>You've reverted yourself...</b>
Please start	<b>Please do not remove warnings...</b>
Direct question	<b>What is your native language?</b>
Negative lexicon	<b>If you're going to accuse me...</b>

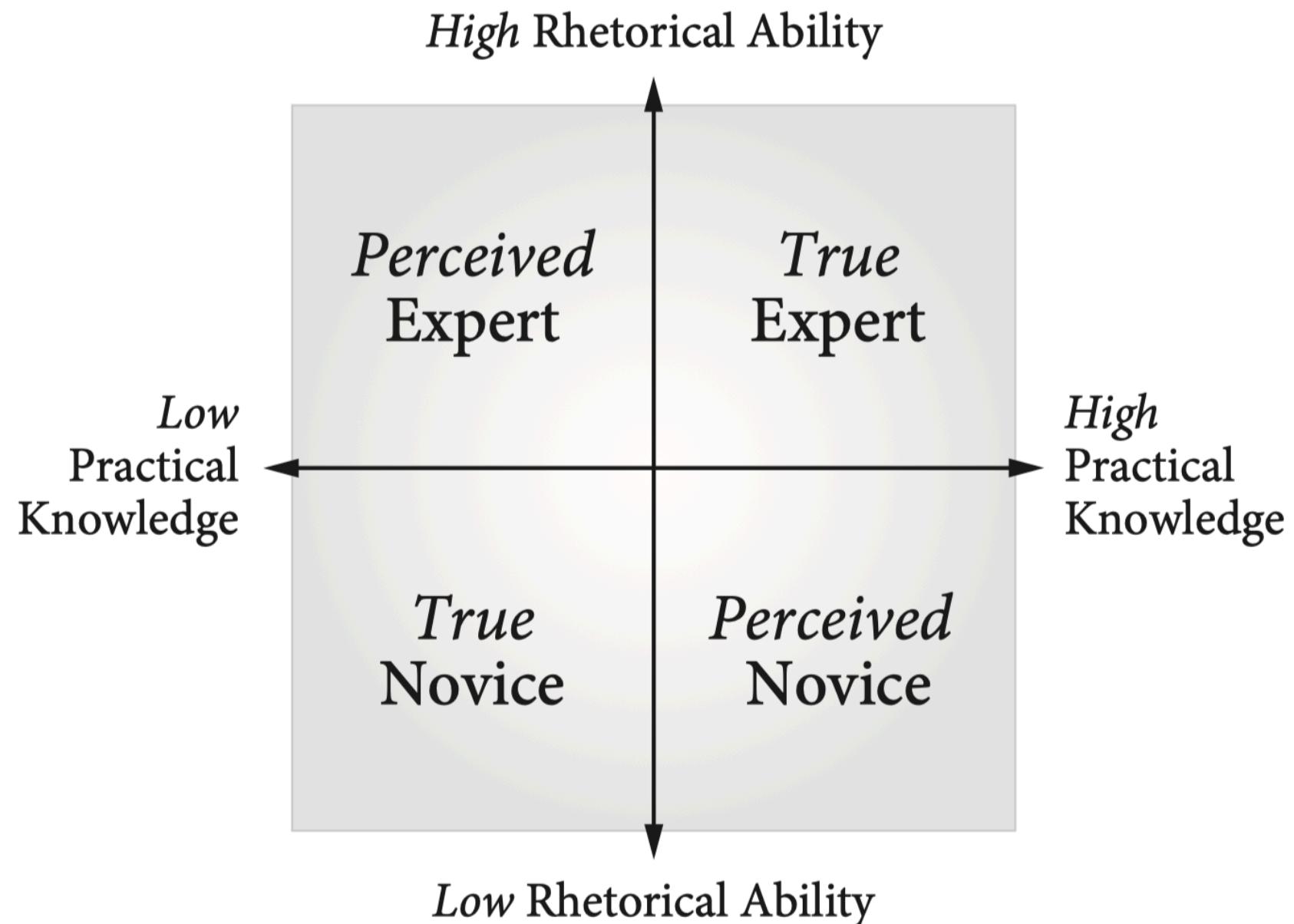
<sup>7</sup> Danescu-Niculescu-Mizil et al., 2013, A computational approach to politeness with application to social factors

# Expressing Expertise<sup>8</sup>

**Definition:** Convincing users that the agent is an expert on its subject matter.  
Expertise has two dimensions:

**Rhetorical ability:** speaking prowess.

**Practical knowledge:** prior knowledge and experience on the topic.



<sup>8</sup> Andrist et al., 2013, Rhetorical robots

## Expertise Cues<sup>8</sup>

---

### Goodwill

*Wanting the best for the listener.*

## Examples

---

**Expert:** “This cafe is a great place to go for lunch to get out of the hot sun.”

**Novice:** “This cafe is a great place to go for lunch.”

---

### Prior expertise

*References to past helping experience.*

**Expert:** “I send a lot of visitors to this museum each year.”

**Novice:** “A lot of visitors go to this museum each year.”

---

### Organization

*More natural organization of information.*

**Expert:** “At 1000 years old, the castle is the the oldest landmark in the city. It has Gothic architecture.”

**Novice:** “The castle is 1000 years old. It has Gothic architecture. It’s the oldest landmark in the city.”

---

### Metaphors

*Making descriptions more accessible.*

**Expert:** “Stepping onto the sunny beach is like wrapping yourself in a towel from the dryer.”

**Novice:** “The sunny beach is quite hot.”

---

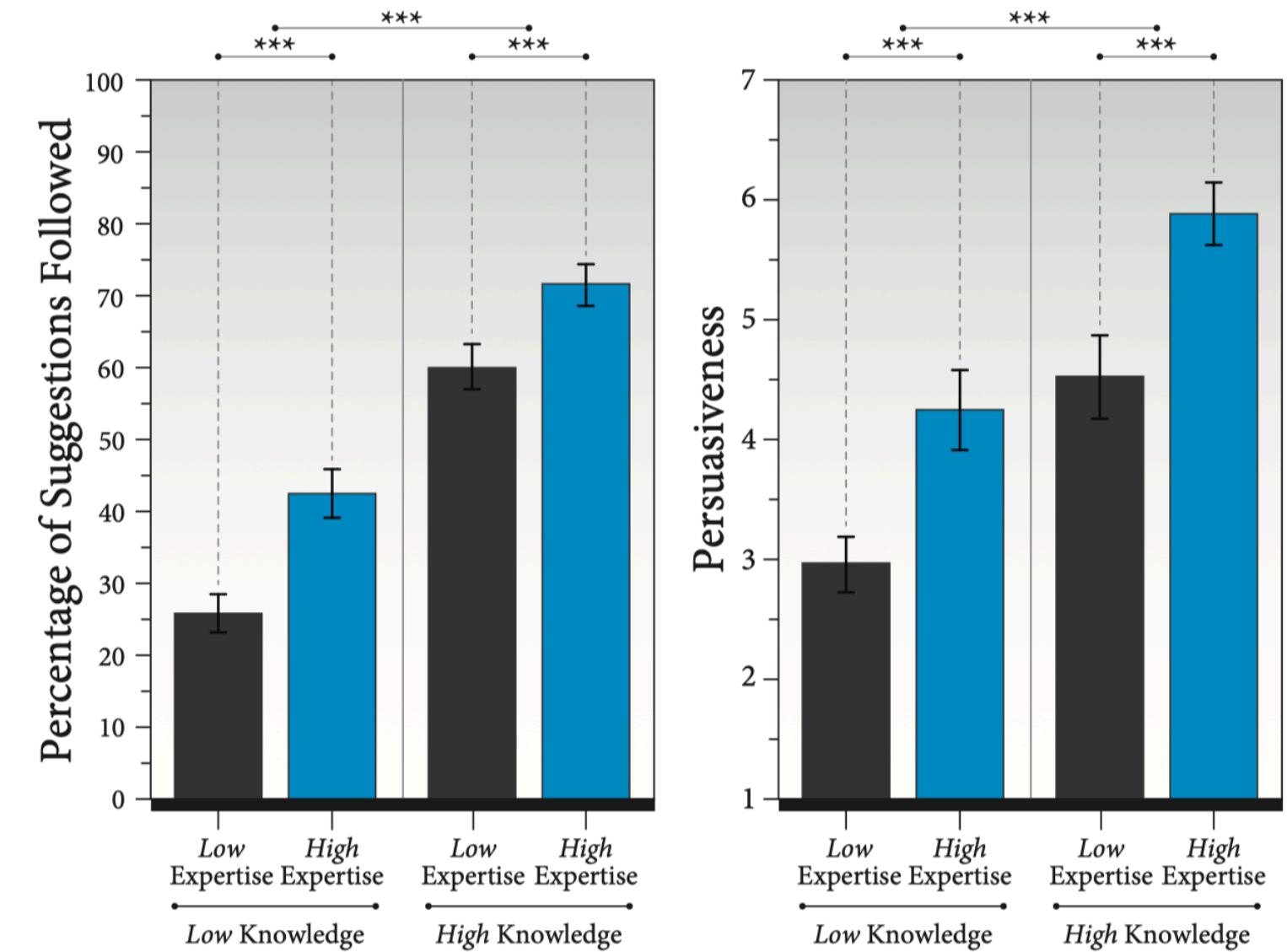
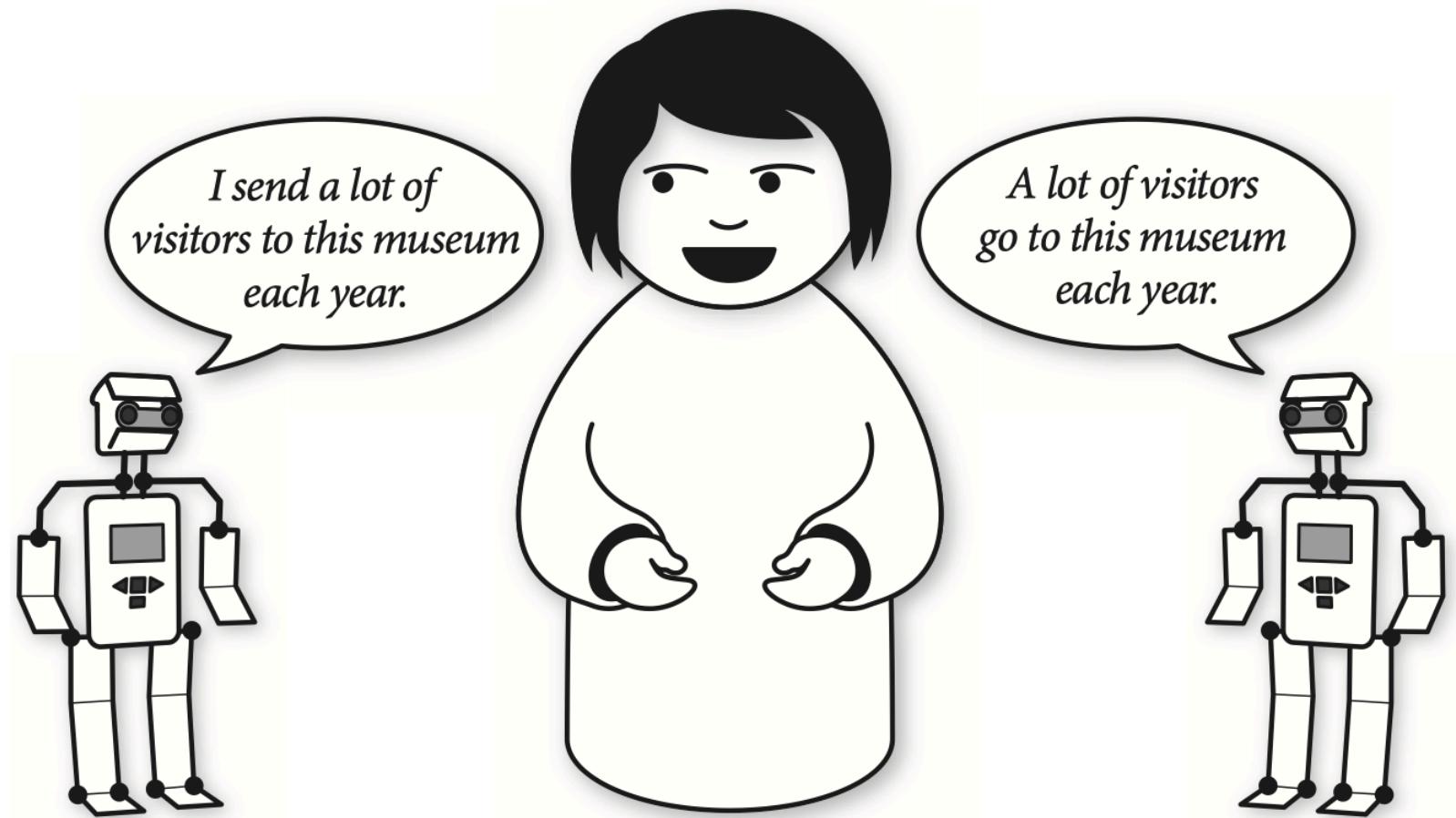
### Fluency

*Reduced pauses and confidence in speech.*

**Expert:** “The statue is 200 years old. [A 300 ms. pause] It was built to honor the King.”

**Novice:** “The statue is 200 years old. [A 1200 ms. pause] It was built to honor the King.”

<sup>8</sup> Andrist et al., 2013, Rhetorical robots



# Personality in Artificial Agents

## What is personality?

**Definition:** *Personality* refers to individual differences in characteristic patterns of thinking, feeling, and behaving.<sup>9</sup>

There are two prevailing models of personality:

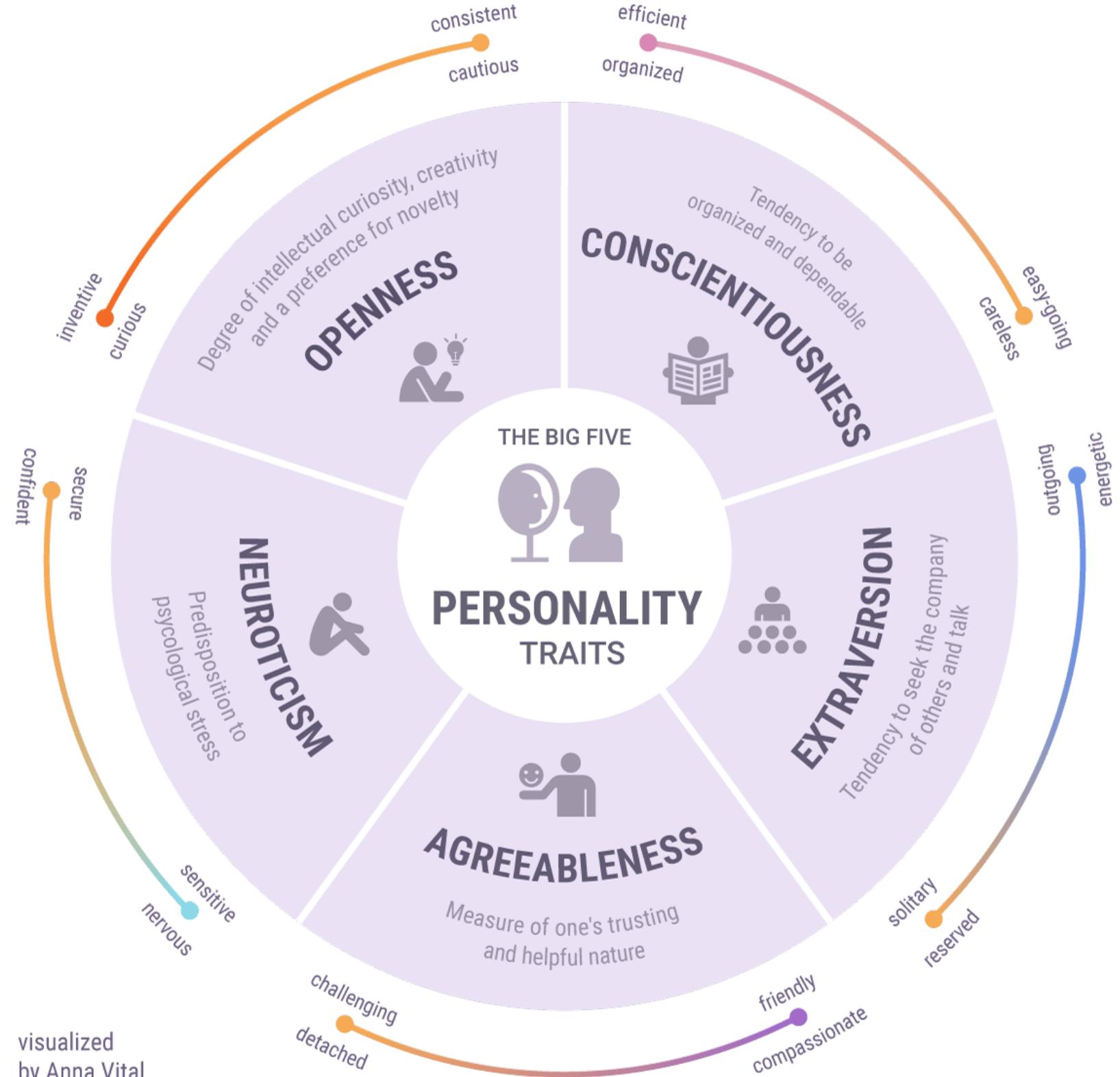
1. Personality traits
2. Personality types

<sup>9</sup> APA

# Personality Traits<sup>10 11</sup>

The most commonly accepted set of traits are the Big Five:

1. Openness
2. Conscientiousness
3. Extraversion
4. Agreeableness
5. Neuroticism



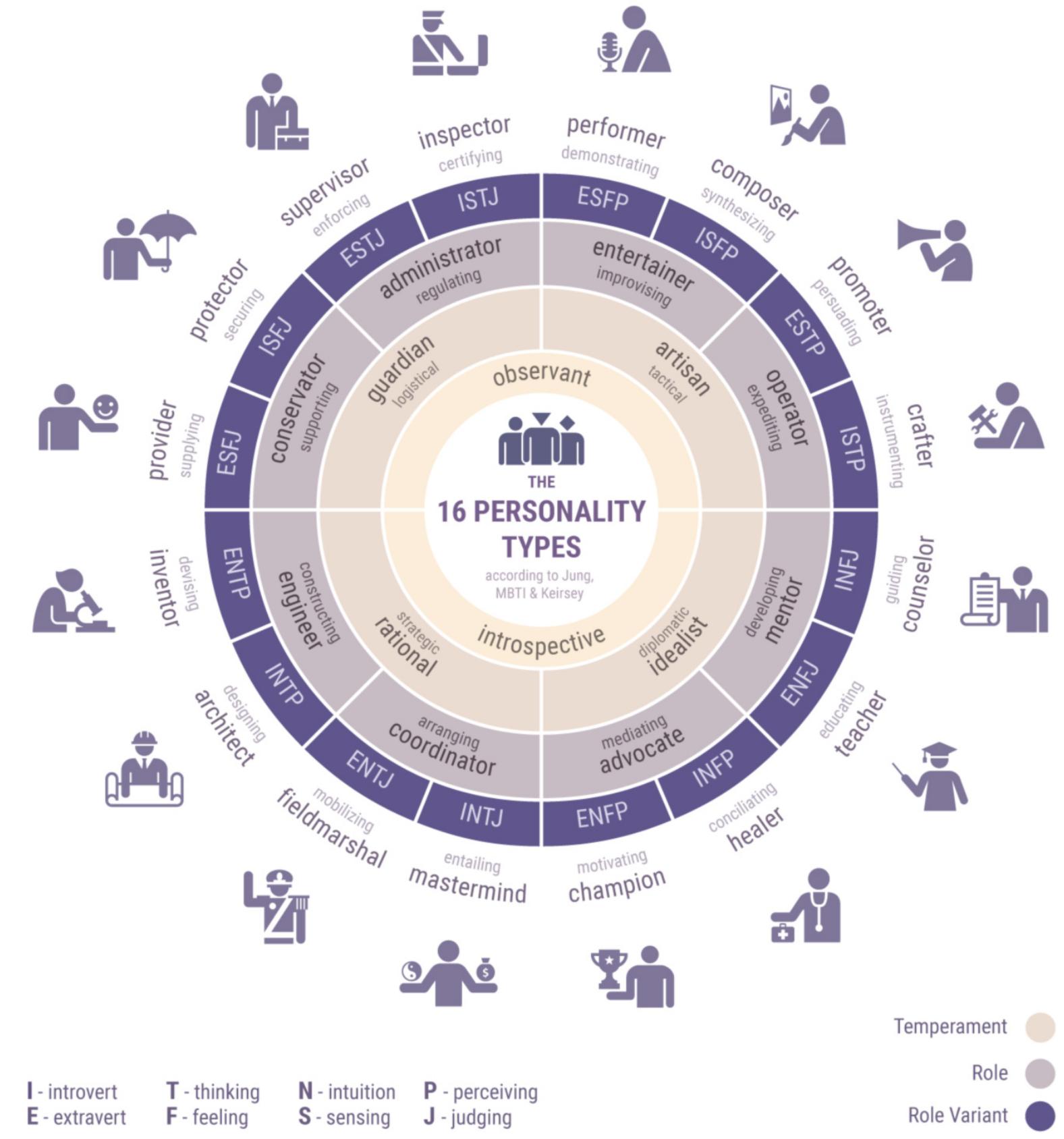
<sup>10</sup> [Image source](#)

<sup>11</sup> [Live Science](#)

# Personality Types<sup>12</sup>

The Myers-Briggs theory, the most commonly accepted set of personality types, posits that there are 16 distinct personalities that vary across four dimensions:

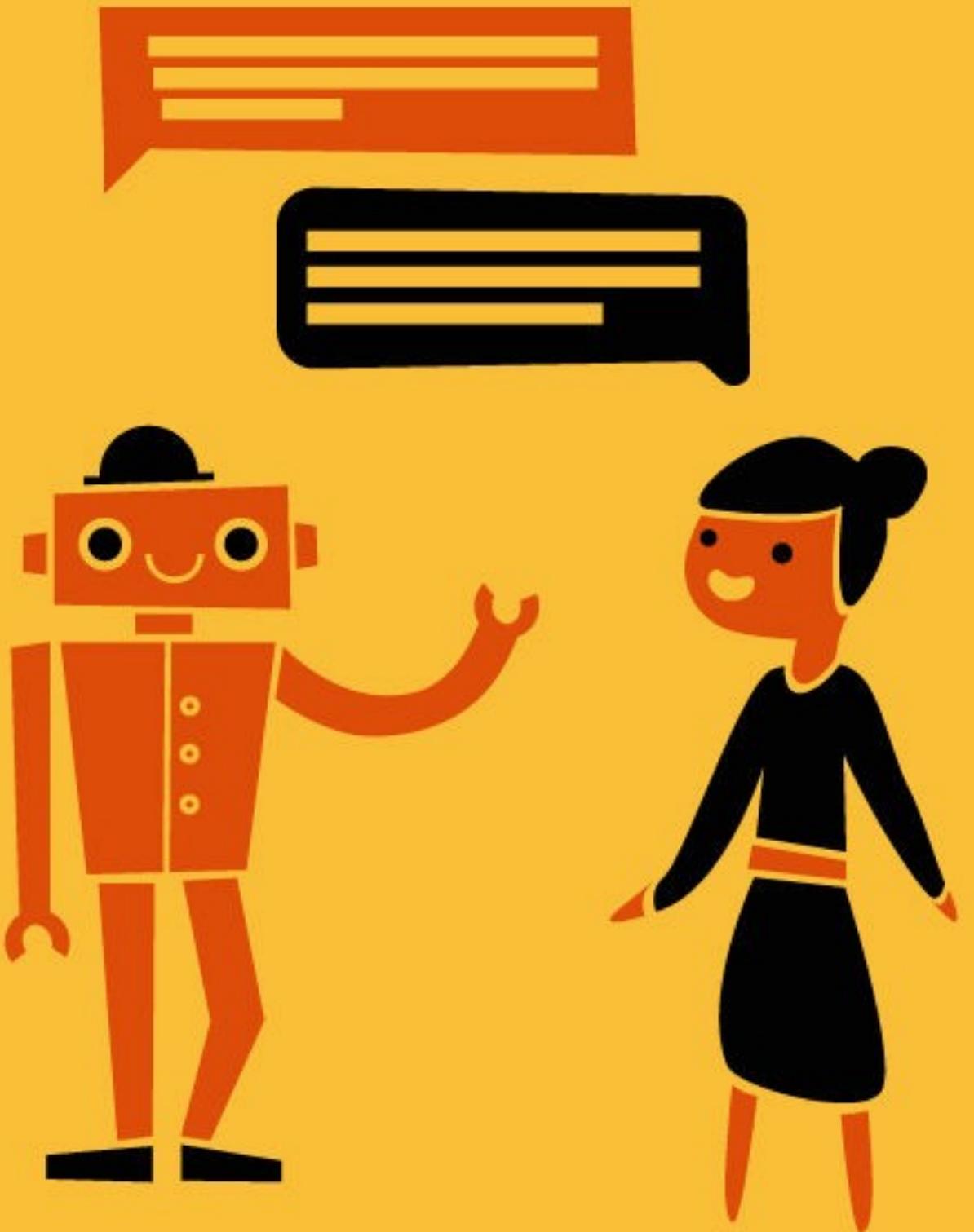
1. Extraversion and introversion
2. Sensing and intuition
3. Thinking and feeling
4. Judgment and perception



<sup>12</sup> [Image source](#)

# Personality Design Strategies<sup>13</sup>

- Personality matching
- Personality expression
- Persona development



<sup>13</sup> [Image source](#)

## Personality Matching

There are two methods for matching the personality of the agent with the personality of the user:

1. The agent and the user have the *same* personality

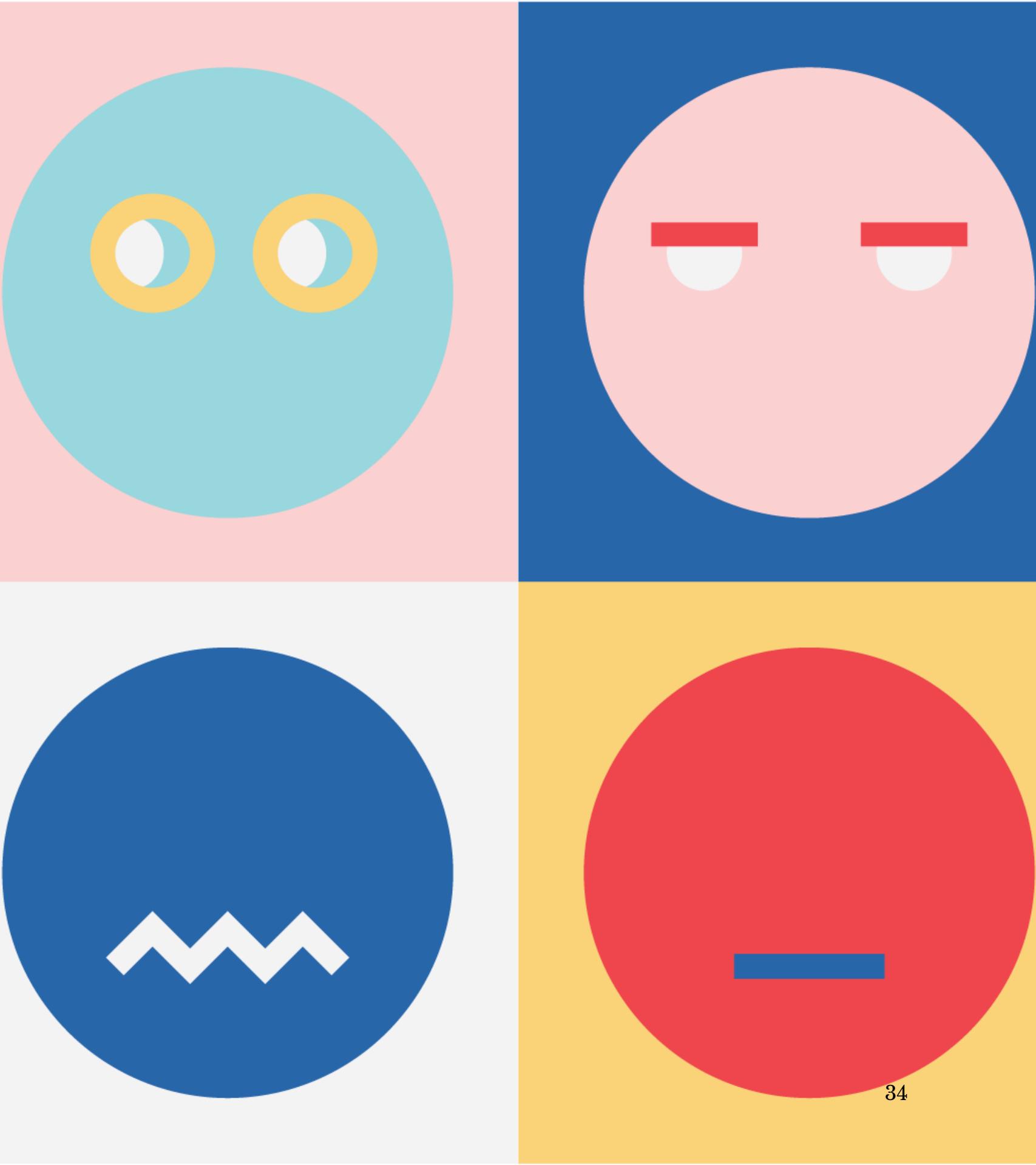
**Similarity-attraction theory would suggest that users will favor agents that have the same personality .**

2. The agent and the user have *complementary* personalities

E.g., an agent can be designed to express the ISTP (crafter) personality to work with an INTJ (mastermind) user.

## Further Reading<sup>18</sup>

- The conversation designer's handbook
- A guide to developing bot personalities
- What are & how to create personas
- A closer look at personas



<sup>18</sup> Image source

# What did we learn today?

- How can we understand computers as social actors?
- How can we design character speech?
- What is personality in artificial agents?

# Building User Interfaces

# Usability Evaluation

Cole Nelson

# What will we learn today?

- Why do we evaluate?
- How can we redefine usability?
- How can we measure usability?
- What are the usability testing basics?
- How can we design a user test?

# Why evaluate?

## Recap: What is UX design?<sup>19</sup>

**Definitions:** User experience (UX) design is the process design teams use to create products that provide meaningful and relevant experiences to users.

<sup>19</sup> Source: [Interaction Design Foundation](#)

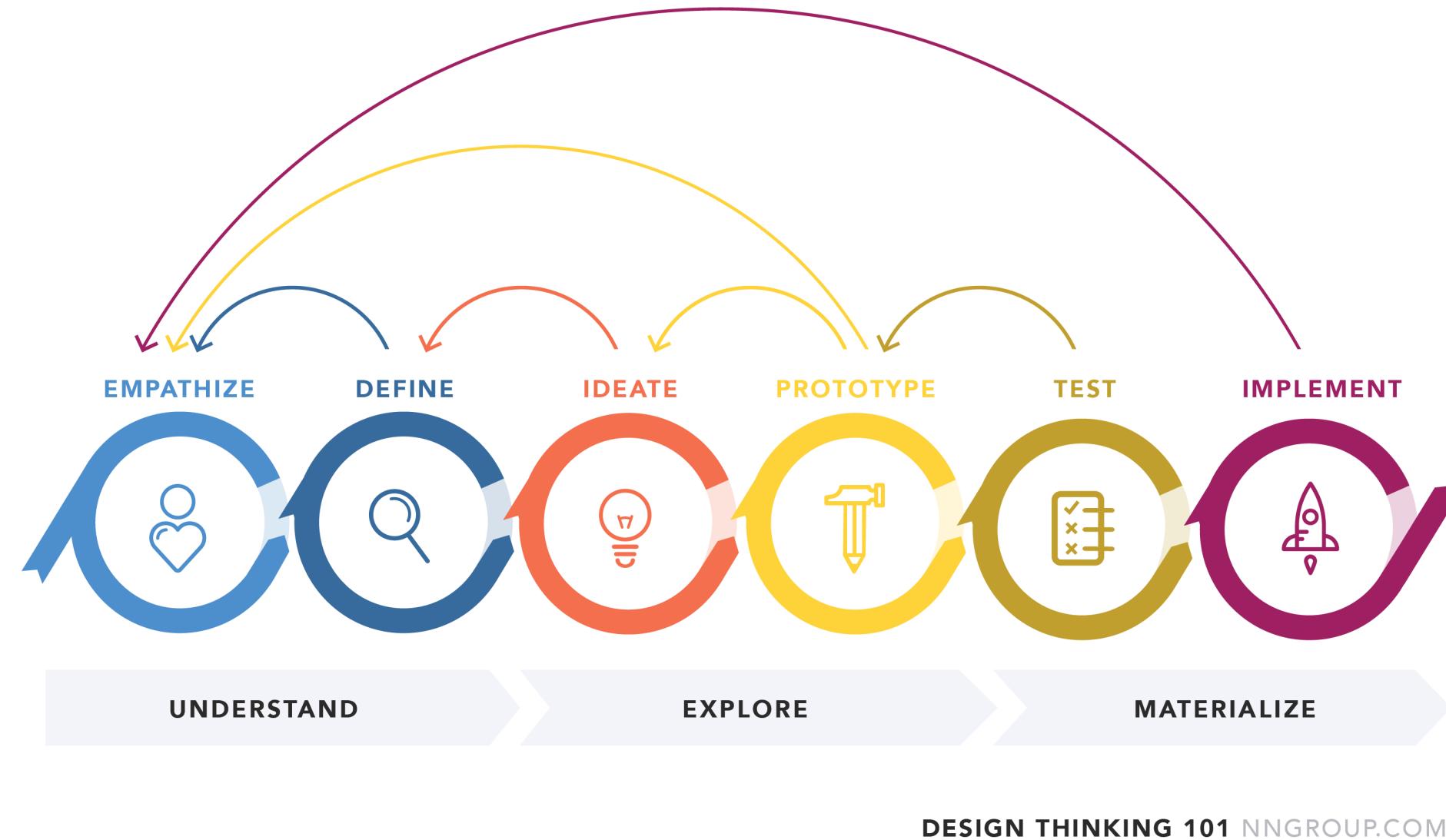
## Recap: What is the design process?<sup>20</sup>

UX design usually involves the steps:

1. Empathize
2. Define
3. Ideate
4. Prototype
5. Test
6. Implement



<sup>20</sup> Image source: [NN/g Design Thinking](#)



<sup>21</sup> Image source: [NN/g Design Thinking](#)

## Recap: Usability Evaluation

**Usability:** The *effectiveness*, *efficiency*, and *satisfaction* with which a specified set of users can achieve a specified set of tasks in a particular environment. — ISO 9241-11

**Usability Evaluation:** The assessment of the usability of design solutions.

# Redefining Usability

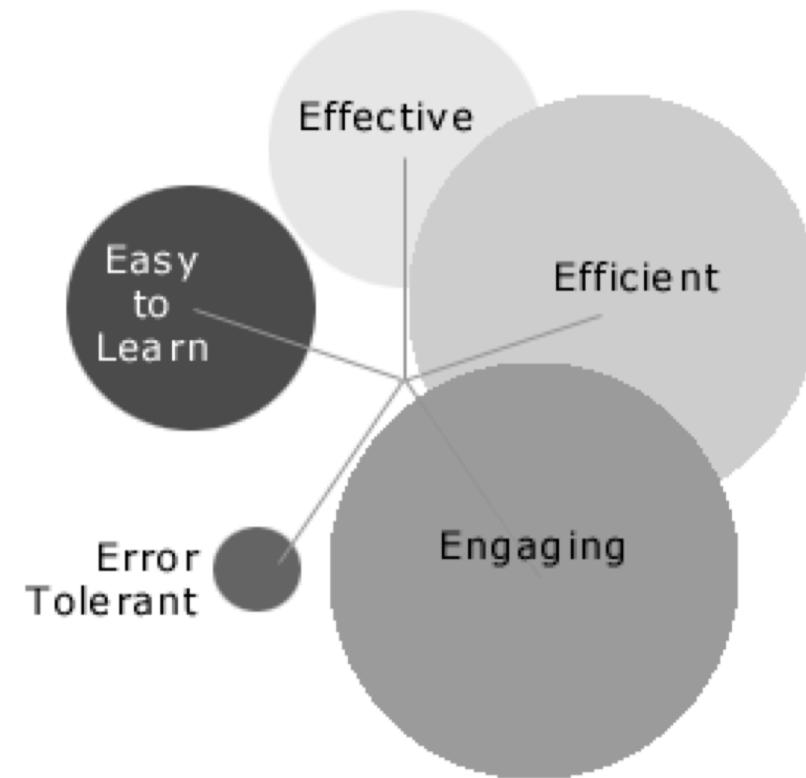
# The Five-E Model of Usability<sup>22</sup>

Dimension	Definition
Effective	How completely and accurately the work or experience is completed or goals reached
Efficient	How quickly this work can be completed
Engaging	How well the interface draws the user into the interaction and how pleasant and satisfying it is to use
Error tolerant	How well the product prevents errors and can help the user recover from mistakes that do occur
Easy to learn	How well the product supports both the initial orientation and continued learning throughout the complete lifetime of use

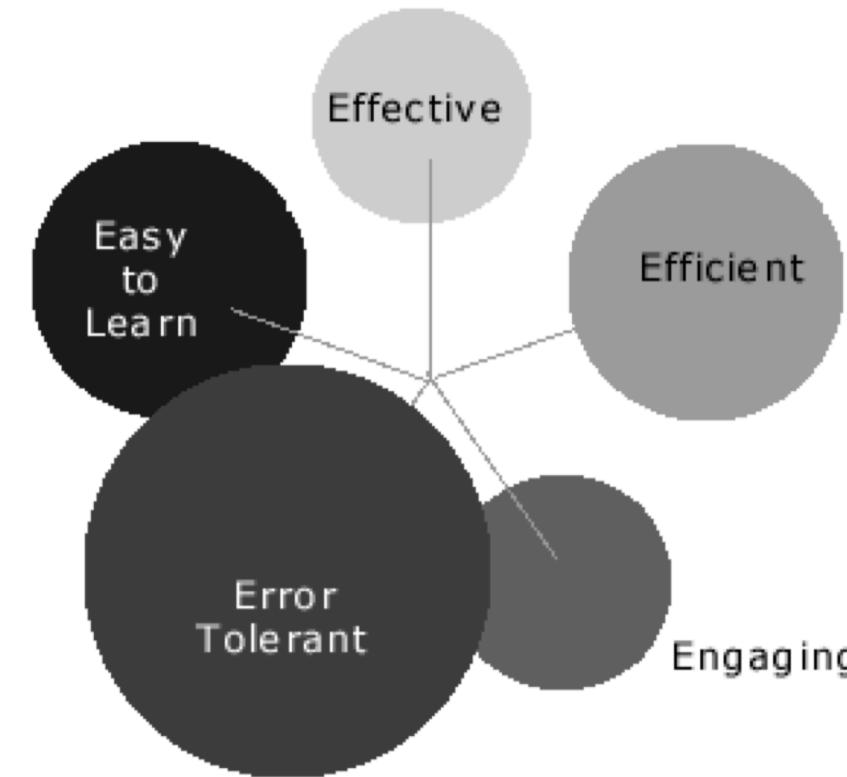
<sup>22</sup> Quesenberry, 2003, Dimensions of Usability

Different design and evaluation problems will require considering the five Es with different weights.<sup>23</sup>

Five Es for a *museum site*:



Five Es for a *registration form*:



<sup>23</sup> Image source: [Quesenberry, 2003, Dimensions of Usability](#)

# Measurement

# Types of Measures

1. Performance metrics
2. Issue-based metrics
3. Self-report metrics

## Performance Metrics

**Task success:** measures how effectively users are able to complete a given set of tasks. Can be used as binary or levels of success.

**Time-on-task** measures how much time is required to complete a task.

**Errors** measures the mistakes made during a task.

**Efficiency** measures the level of effort required to complete the task.

**Learnability** measures how performance changes over time.

## Issue-based Metrics<sup>24</sup>

**Definition:** Problems that users encounter in using a system.

Examples:

- Behaviors that prevent task completion
- Behaviors that takes someone “off course”
- An expression of frustration by the participant
- Not seeing something that should be noticed

<sup>24</sup> Albert & Tullis, 2013, Measuring the User Experience

- Participant says a task is complete when it is not
- Performing an action that leads away from task success
- Misinterpreting some piece of content
- Choosing the wrong link to navigate through web pages

## **Self-report metrics**

**Definition:** Asking participants about their perceptions of and experience with a design solution using a set of questions.

Participants provide quantitative (e.g., ratings, rankings) or qualitative (e.g., open-ended, narrative) responses.

Commonly used self-report metrics include SUS and USE.

**SUS<sup>25 26</sup>**

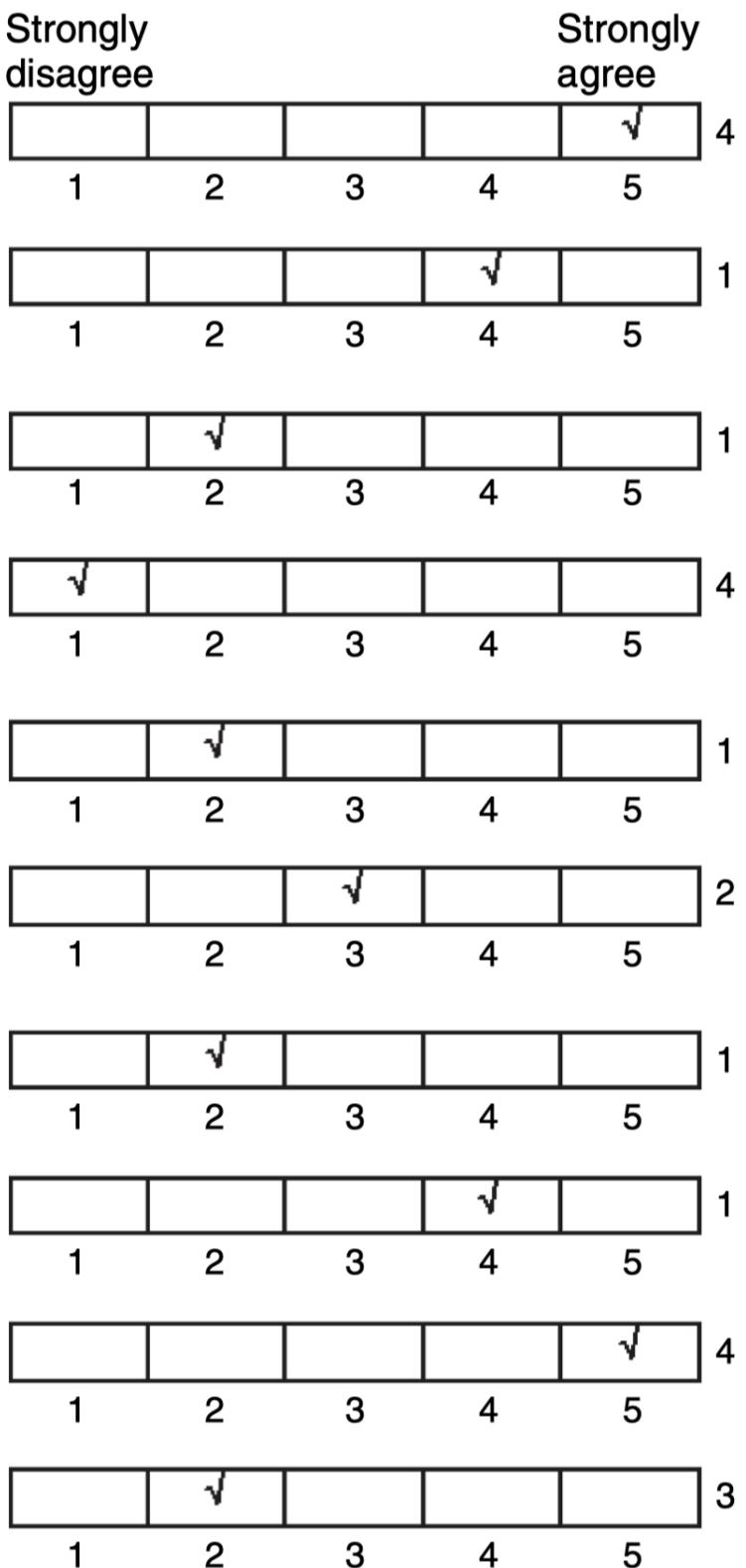
Ten-item questionnaire that focuses on usability.

Can be used for relative comparison or absolute benchmarking.

<sup>25</sup> How to use the SUS

<sup>26</sup> Image source: Albert & Tullis, 2013, Measuring the User Experience

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought this system was too inconsistent.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.



**Total = 22**

**SUS Score =  $22 \times 2.5 = 55$**

Includes four sub-scales for usefulness, ease of use, ease of learning, and satisfaction.

**Usefulness**

- It helps me be more effective.
- It helps me be more productive.
- It is useful.
- It gives me more control over the activities in my life.
- It makes the things I want to accomplish easier to get done.
- It saves me time when I use it.
- *It meets my needs.*
- It does everything I would expect it to do.

**Ease of Use**

- It is easy to use.
- It is simple to use.
- It is user friendly.
- It requires the fewest steps possible to accomplish what I want to do with it.
- *It is flexible.*
- *Using it is effortless.*
- *I can use it without written instructions.*
- *I don't notice any inconsistencies as I use it.*
- *Both occasional and regular users would like it.*
- *I can recover from mistakes quickly and easily.*
- *I can use it successfully every time.*

**Ease of Learning**

- I learned to use it quickly.
- I easily remember how to use it.
- It is easy to learn to use it.
- *I quickly became skillful with it.*

**Satisfaction**

- I am satisfied with it.
- I would recommend it to a friend.
- It is fun to use.
- It works the way I want it to work.
- It is wonderful.
- I feel I need to have it.
- It is pleasant to use.

Users rate agreement with these statements on a 7-point Likert scale, ranging from strongly disagree to strongly agree. Statements in *italics* were found to weight less heavily than the others.

<sup>27</sup> Image source: [Albert & Tullis, 2013, Measuring the User Experience](#)

## Likert Scales

A numerical 3-5-7-9-11 point scale with descriptive labels. E.g.:

1. Strongly disagree
2. Disagree
3. Neither agree nor disagree
4. Agree
5. Strongly agree

## **Recap: Types of Usability Evaluation**

- 1. User-testing-based methods**
- 2. Expert-review-based methods**

## Recap: User-testing-based methods

**Definition:** Empirical, i.e., based on data, testing with users who represent the target population of design solutions.

Today, we will cover user-testing-based methods.

# Usability Evaluation Basics

# Usability Testing

**Definition:** Observing users performing tasks with a design solution and asking them questions about their experience with the solution.

Observations include user actions, behavior, and verbal descriptions.

## When do we use usability testing?

Depending on where usability testing is used in the design process, the testing can take two forms:

1. *Formative* testing
2. *Summative* testing

## Formative Testing

**Definition:** Testing done throughout the design process to diagnose and address design problems.

Involves small number of users; used repeatedly; informs design improvements.

It "forms" the next iteration of the design.

## How small can the testing be?<sup>28</sup>

Formative testing is considered to be a "discount" usability method, as 85% of usability solutions can be identified through testing with 5 users.

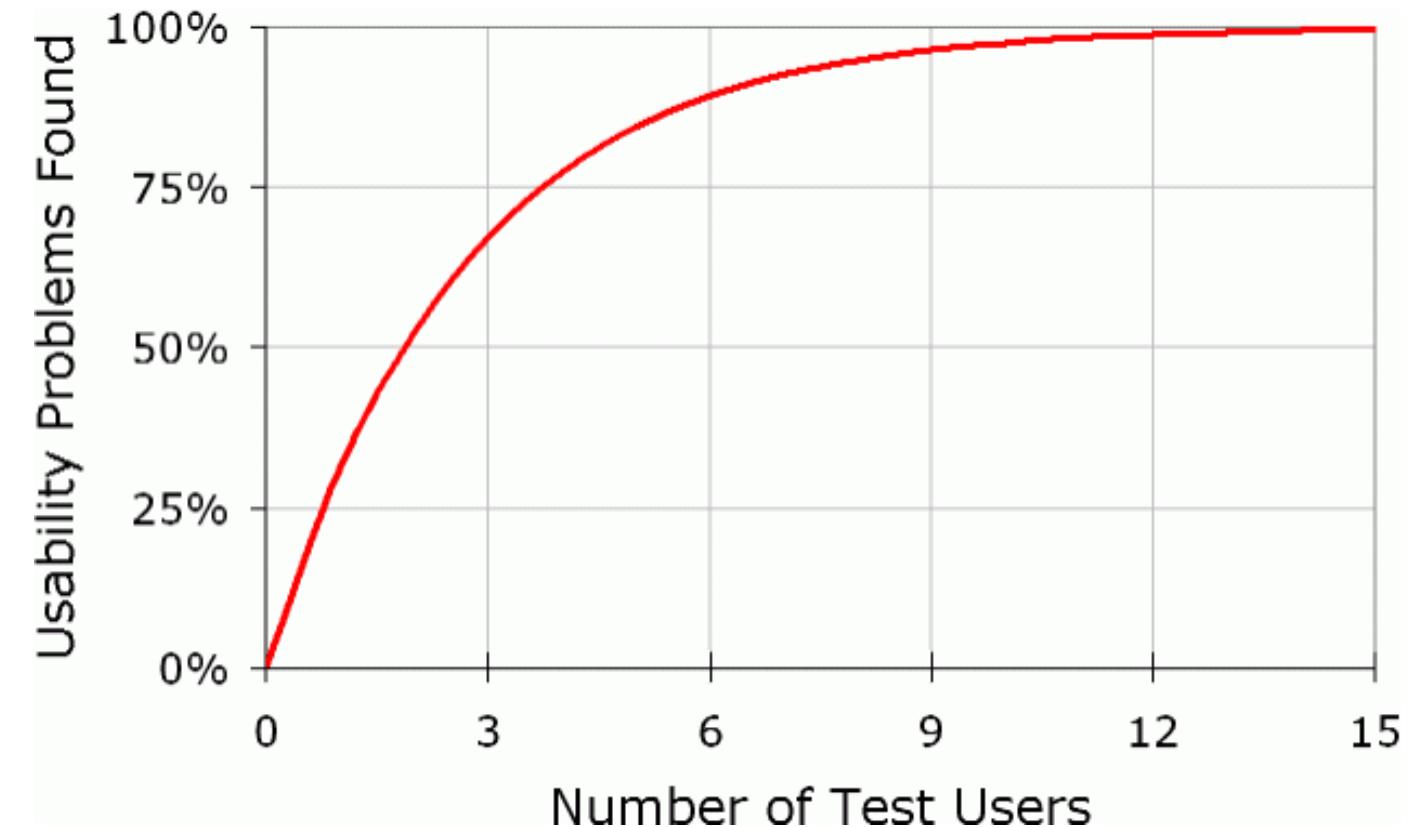
$$\text{Number of problems found} = N(1 - (1 - L)^n)$$

$N$ : total number of usability problems

$n$ : number of users

$L$ : the proportion of usability problems discovered while testing a single user

( $L = 31\%$ )



<sup>28</sup> Image source: [NN/g: Why you only need to test with 5 users](#)

## Summative Testing

**Definition:** Testing done at the end of the design process to establish the baseline usability of the design solution.

Involves a larger number of users; comparative testing; utilizes large number of metrics and statistics methods.

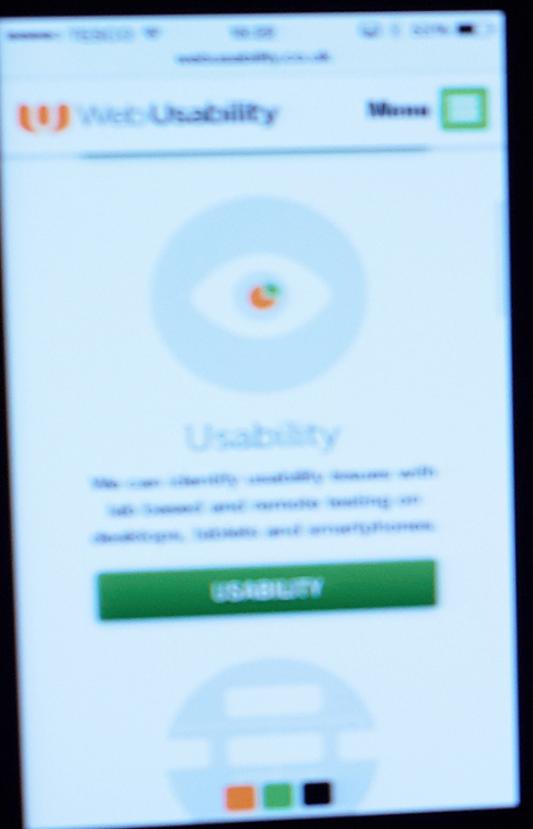
# Usability Testing Contexts

- *Laboratory* testing
- *Field* testing
  - "Guerilla" testing
  - Remote testing

## Laboratory Testing<sup>29</sup>

**Definition:** Testing in the lab set up to capture user behavior through screen recording, software logging, over-the-shoulder video recording, eye tracking, etc. and to allow the design team to observe and analyze the test session.

<sup>29</sup> [Image source](#)





<sup>30</sup> Image source

## Field Testing

**Definition:** Testing in the target setting of use for the design solution with the target profile of users.

## Field Methods: Guerilla Testing<sup>31</sup>

**Definition:** Low-cost usability testing set up in a public space where passersby are recruited as test participants as volunteers or small compensation.



<sup>31</sup> [Image source](#)

## **Field Methods: Remote Testing**

**Definition:** Testing a hi-fi prototype or early version of a deployed product over the internet.

Different forms of remote testing:

- **Moderated:** expert guides and observes, asks questions
- **Unmoderated:** participants completes tasks on their own schedule, captures behavior (e.g., A/B testing)

# Designing a Usability Test

# **Key Dimensions of Usability Testing**

When designing a usability test, we need to define and characterize the following four dimensions:

**Why**

Goals

**What**

Scope, task/  
scenarios

**How**

Approach,  
metrics

**Who**

User  
subgroups,  
study team

## The Why

Defining the why will involve determining test goals and provide a 10,000-feet view of the design goals.

- E.g., "improving accessibility of a website for older adults"

Different goals will result in entirely different test designs.

## Formulating Test Goals

Formulate goals as questions that the test is designed to answer and specify two components.

**Does our solution significantly improve accessibility for older adults over the previous design?**

Points toward a comparative test with older adults.

**To what extent do users consider our solution to be usable?**

Points toward a study with standard metrics.

Test goals should specify:

1. **Desired outcomes** capture how the design is expected to achieve. E.g., improved accessibility, reduced error rate, etc.
2. **Basis for comparison** specifies whether the outcome is with respect to a baseline, such as a previous design, established guidelines, or performance expectations. E.g., minimum score on a standardized test.

## The What

We need to determine the scope of the testing, including what aspects of the system design, what tasks, and what scenarios will be included in the testing.

- Are we looking to evaluate the *breadth* of functionality?  
→ **horizontal prototype**
- Are we looking to evaluate the *depth* of functionality?  
→ **vertical prototype**

## Developing the **What**

Defining the *what* involves defining:

- **Questions:** Expectations of specific outcomes, e.g., whether or not the user will successfully achieve a particular goal.
- **Tasks:** The sequence of actions that users are expected to perform to achieve goals.
- **Scenarios:** Brief stories that provide users with context and goals in using the system.

## An Example<sup>32</sup>

**Question:** Will users look at the top navigation bar to start their search for information?

**Task:** Seeking information about online programs for military personnel. Correct choice is Featured Degrees in top navigation bar. Users can also find a link to programs for military personnel in the description of featured programs in the center of the homepage, but it may be below the fold on their computer screen.

<sup>32</sup> Barnum, 2011, Usability Testing Essentials

**Scenario:** You have a friend in the military who wants to enroll in college courses while serving. You want to see if there are any online programs your friend could apply for. How would you go about doing this on this website?

## How do we present multiple scenarios?

Scenarios should be ordered:

- From initial impressions to specific tasks
- From general to specific
- From short to long
- From simple to complex

Scenarios can also be presented *all at once or one at a time* depending on the testing goals.

## The **How**

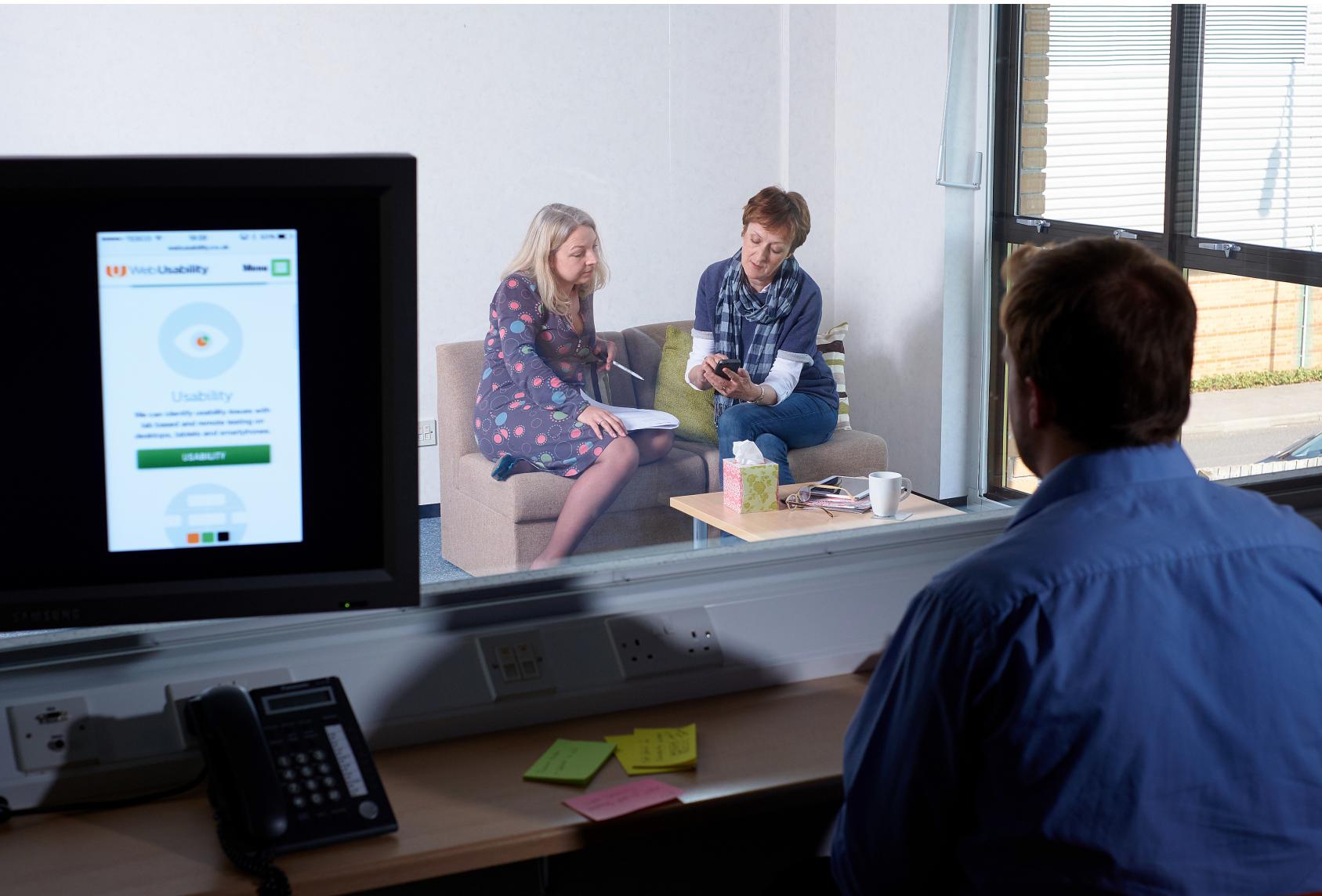
The how of the usability test will depend on:

1. Whether the test is *formative* or *summative*
2. Whether the test is for a *single design* or *comparative*

# Types of Measurement<sup>33</sup>

We can collect two types of data:

1. **Qualitative data:** observations of user actions and behavior, comments, and answers to questions
2. **Quantitative data:** measurements of user performance, error, and perceptions of the design



<sup>33</sup>[Image source](#)

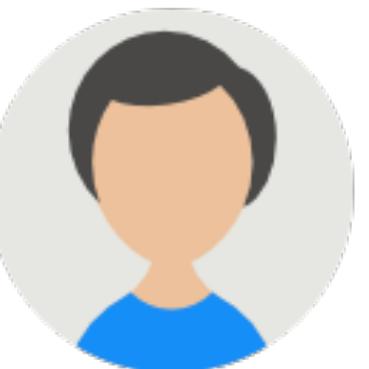
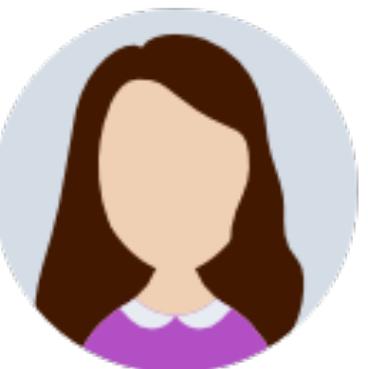
## The Who

The *who* of the usability test includes:

1. Participants who represent the target user population
2. Team roles during usability testing

# Participants

Participants should represent the user subgroups that is targeted by the design. Subgroups characteristics can be defined by *experience, familiarity, skill, occupation, domain knowledge, and demographics*.



Once user subgroups are identified, several sessions of the study can be planned for each or a subset of the subgroups. The participants should be representative of the targeted subgroups.

The problem domain should also dictate participant representation. E.g., in a test for a budgeting app, users from different income levels can provide different insights.

## Team Roles

The testing team usually involves:

1. **Moderator** who guides the participant and probes them with questions.
2. **Note-taker** who captures data.
3. **Observer(s)** from the UX team.
4. **Technician**, who operates the tested system or the testing equipment.

## The Test Plan

The outcome of the design of the usability test is a *test plan* document that captures the *why, what, how, and who*.

Sometimes called a *test protocol*.

Supplements can include checklists for each role, moderator script, consent form, non-disclosure agreement (NDA) form.

# What did we learn today?

- Why do we evaluate?
- How can we redefine usability?
- How can we measure usability?
- What are the usability testing basics?
- How can we design a user test?