Building User Interfaces

Designing Agents & Characters

Professor Bilge Mutlu

What we will learn today?

— Computers as Social Actors

— Designing Character Speech

— Personality in Artificial Agents

Computers as Social Actors

Computers as Social Actors (CASA)[1]

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.

[1] Nass & Moon, 2000, Machines and Mindlessness: Social Responses to Computers

CASA extends to many social-psychological concepts.[1]

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.

[1] Nass & Moon, 2000, Machines and Mindlessness: Social Responses to Computers

Similarity-Attraction Theory[2]

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

[2] Byrne et al, 1967, Attraction and similarity of personality characteristics.
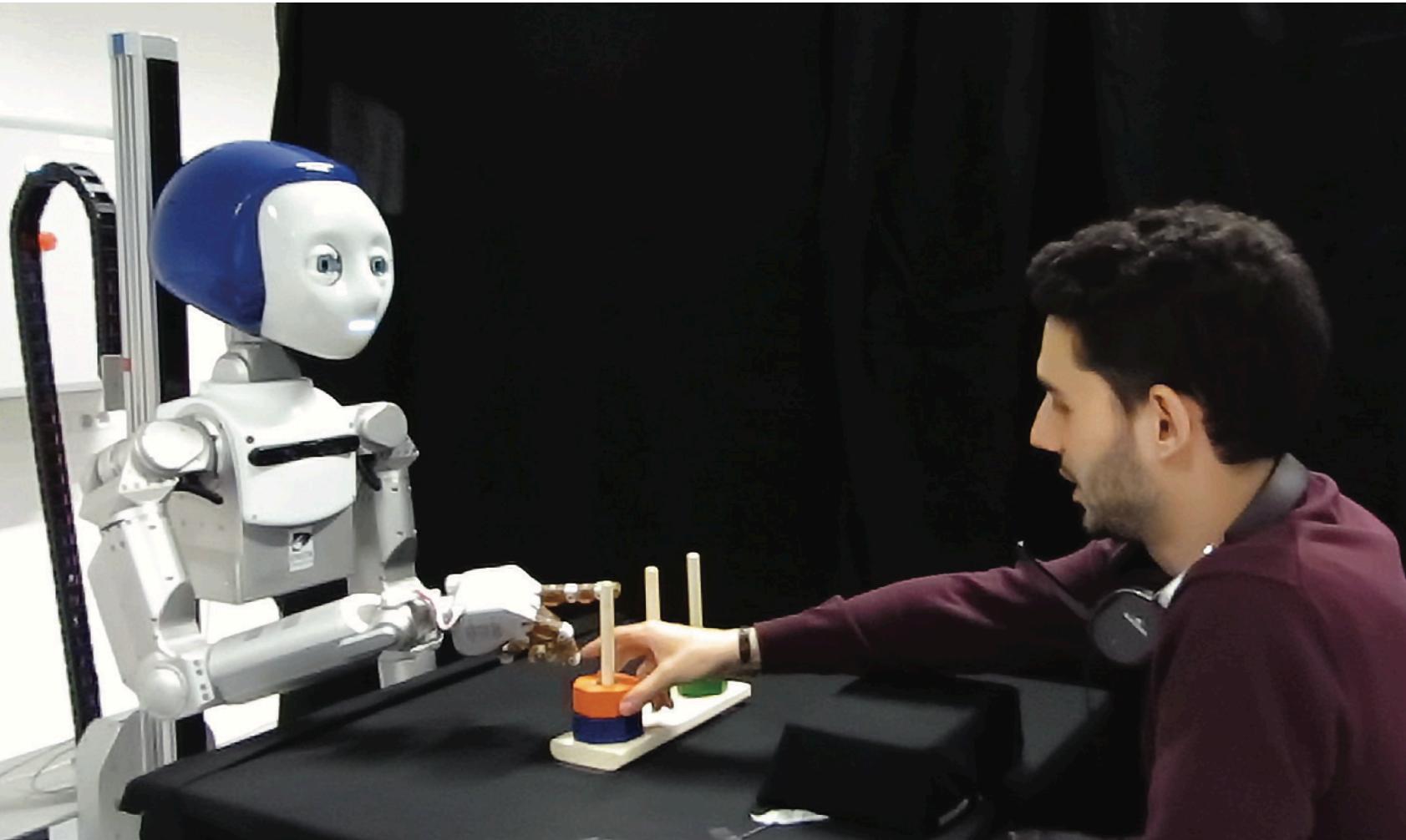
Similarity-Attraction in Character Design

— 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[3]

A robot coach matched the personality of its user purely by increasing/decreasing eye contact.
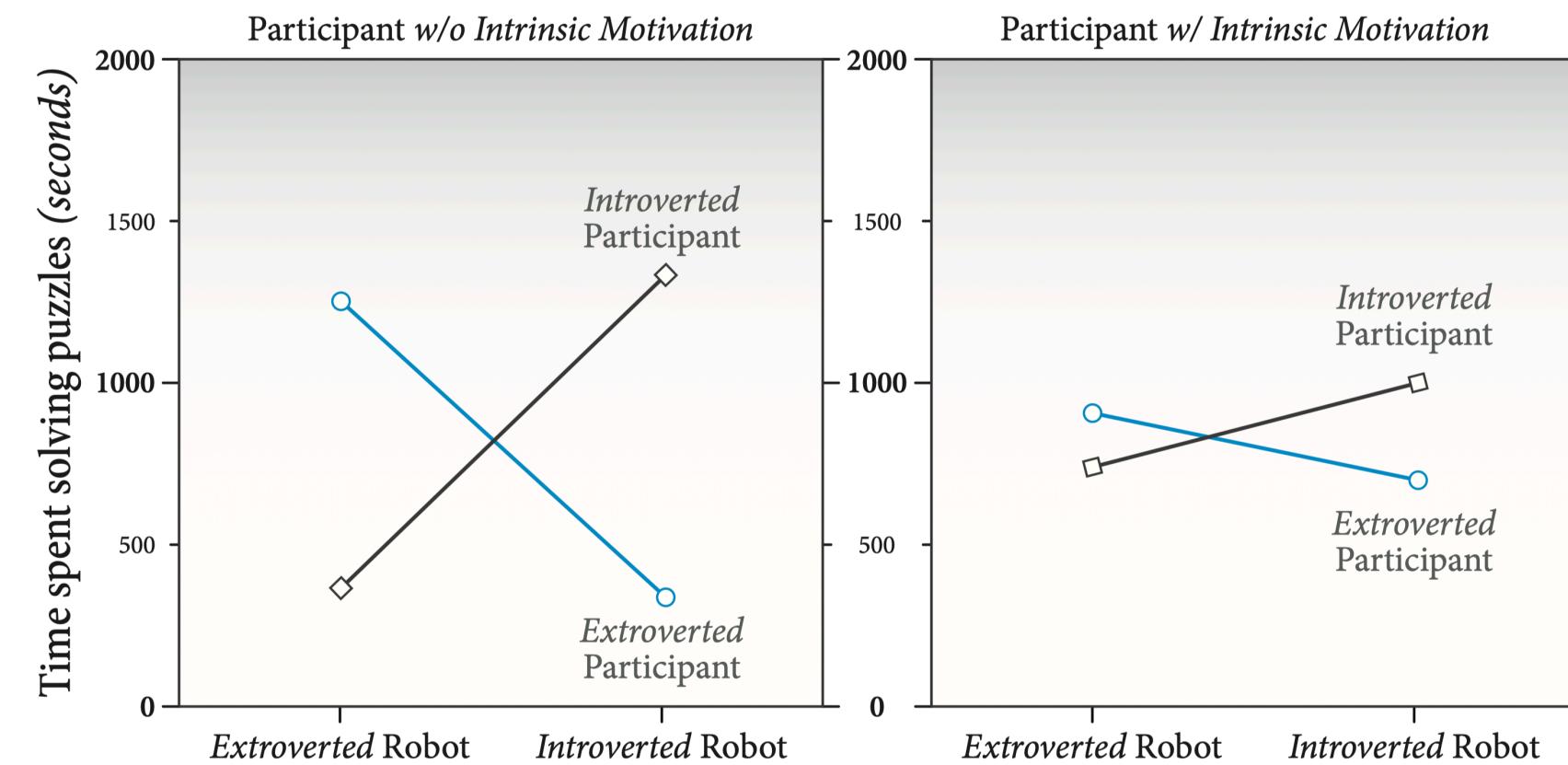
— Extroverts build more eye contact; introverts build less eye contact.

[3] Andrist et al., 2015, Look Like Me...

Results support similarity-attraction theory: people were motivated by a robot with a similar personality.[3]

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

Participant *w/o Intrinsic Motivation*

Introverted Participant

Extroverted Participant

Extroverted Robot   *Introverted* Robot

Participant *w/ Intrinsic Motivation*

Introverted Participant

Extroverted Participant

Extroverted Robot   *Introverted* Robot

Time spent solving puzzles *(seconds)*

[3] Andrist et al., 2015, Look Like Me...

Consistency-Attraction Theory[4]

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.

[4] Nass & Lee, 2001, Does computer-synthesized speech manifest personality?

Consistency-Attraction Example[5]

To create *believable* characters:

— following a dramatic structure with rich backstory and evolving story line

— utilizing verbal and nonverbal social behaviors

— expressing culture

[5] Simmons, 2011, Believable robot characters

Designing Character Speech

Politeness Theory[6]

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

[6] 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 impositon, 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[6]

[6] Brown & Levinson, 1987, Politeness: Some universals in language usage.

102

Positive politeness
Do FTA on record plus
redress to:
H wants [S wants H's wants]

5.3.1 Claim 'common ground'
(S & H ∈ {A} who want {X})

Convey 'X is admirable, interesting'

1. Notice, attend to H (his interests, wants, needs, goods)
2. Exaggerate (interest, approval, sympathy with H)
3. Intensify interest to H

Claim in-group membership with H

4. Use in-group identity markers

Claim common { point of view, opinions, attitudes, knowledge, empathy }

5. Seek agreement
6. Avoid disagreement
7. Presuppose/raise/assert common ground
8. Joke

5.3.2 Convey that S and H are cooperators

Indicate S knows H's wants and is taking them into account

9. Assert or presuppose S's knowledge of and concern for H's wants

Claim reflexivity { If H wants ⟨H has X⟩ then S wants ⟨H has X⟩ / If S wants ⟨S has X⟩ then H wants ⟨S has X⟩ }

10. Offer, promise
11. Be optimistic
12. Include both S and H in the activity
13. Give (or ask for) reasons

Claim reciprocity

14. Assume or assert reciprocity

5.3.3 Fulfil H's want (for some X)

15. Give gifts to H (goods, sympathy, understanding, cooperation)

Fig. 3. Chart of strategies: Positive politeness

Some Examples[6]

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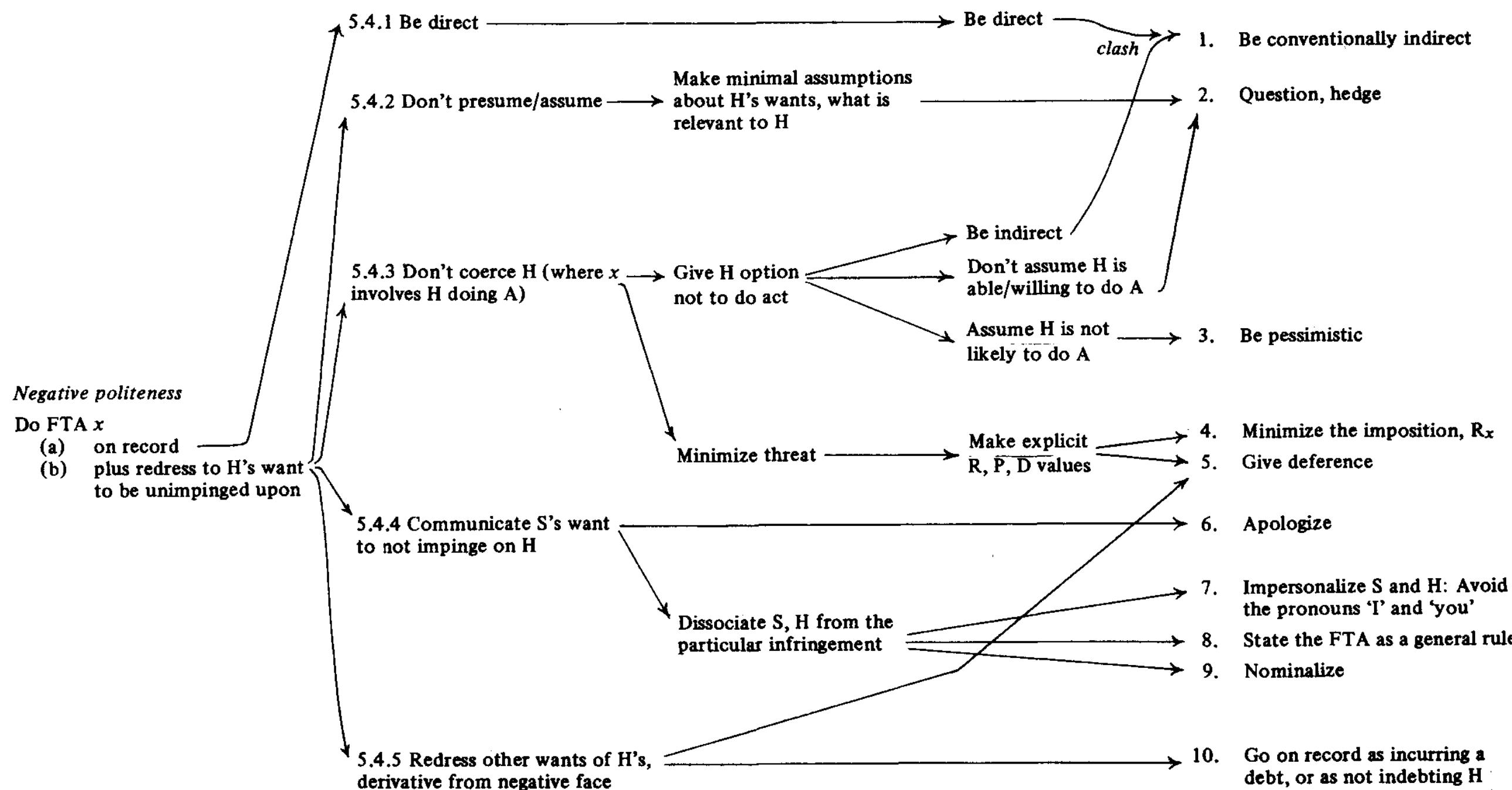
A: I had a flat tyre on the way home.

B: Oh God, a flat tyre!

[6] Brown & Levinson, 1987, Politeness: Some universals in language usage.

Strategies for *Negative* Politeness[6]

---

[6] Brown & Levinson, 1987, Politeness: Some universals in language usage.

Fig. 4. Chart of strategies: Negative politeness

The chart shows:

Negative politeness

Do FTA *x*
(a) on record
(b) plus redress to H's want to be unimpinged upon

5.4.1 Be direct → Be direct → *clash* → 1. Be conventionally indirect

5.4.2 Don't presume/assume → Make minimal assumptions about H's wants, what is relevant to H → 2. Question, hedge

5.4.3 Don't coerce H (where *x* involves H doing A) → Give H option not to do act
- Be indirect
- Don't assume H is able/willing to do A
- Assume H is not likely to do A → 3. Be pessimistic

→ Minimize threat → Make explicit R, P, D values → 4. Minimize the imposition, R*x*
→ 5. Give deference

5.4.4 Communicate S's want to not impinge on H → 6. Apologize
→ Dissociate S, H from the particular infringement → 7. Impersonalize S and H: Avoid the pronouns 'I' and 'you'
→ 8. State the FTA as a general rule
→ 9. Nominalize

5.4.5 Redress other wants of H's, derivative from negative face → 10. Go on record as incurring a debt, or as not indebting H

131
19

Some Examples[6]

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.
You're **quite** right.

Strategy 6: Apologize

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

[6] Brown & Levinson, 1987, Politeness: Some universals in language usage.

Hedging[6]	Example
Less hedging	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?

[6] Brown & Levinson, 1987, Politeness: Some universals in language usage.

See the full list of strategies and examples[6] 👉

---

[6] Brown & Levinson, 1987, Politeness: Some universals in language usage.

Which Strategies Are Most Effective?[7]

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, …

[7] Danescu-Niculescu-Mizil et al., 2013, A computational approach to politeness with application to social factors

Guidelines for Character Design[7]

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 ...

[7] Danescu-Niculescu-Mizil et al., 2013, A computational approach to politeness with application to social factors

Most effective *rudeness* strategies (don't do these):[7]

Strategy	Examples
Direct start	**So** can you retrieve it or not?
Factuality	**In fact** you did link...
2nd person start	**You**'ve reverted yourself...
Please start	**Please** do not remove warnings...
Direct question	**What** is your native language?
Negative lexicon	If you're going to **accuse** me...
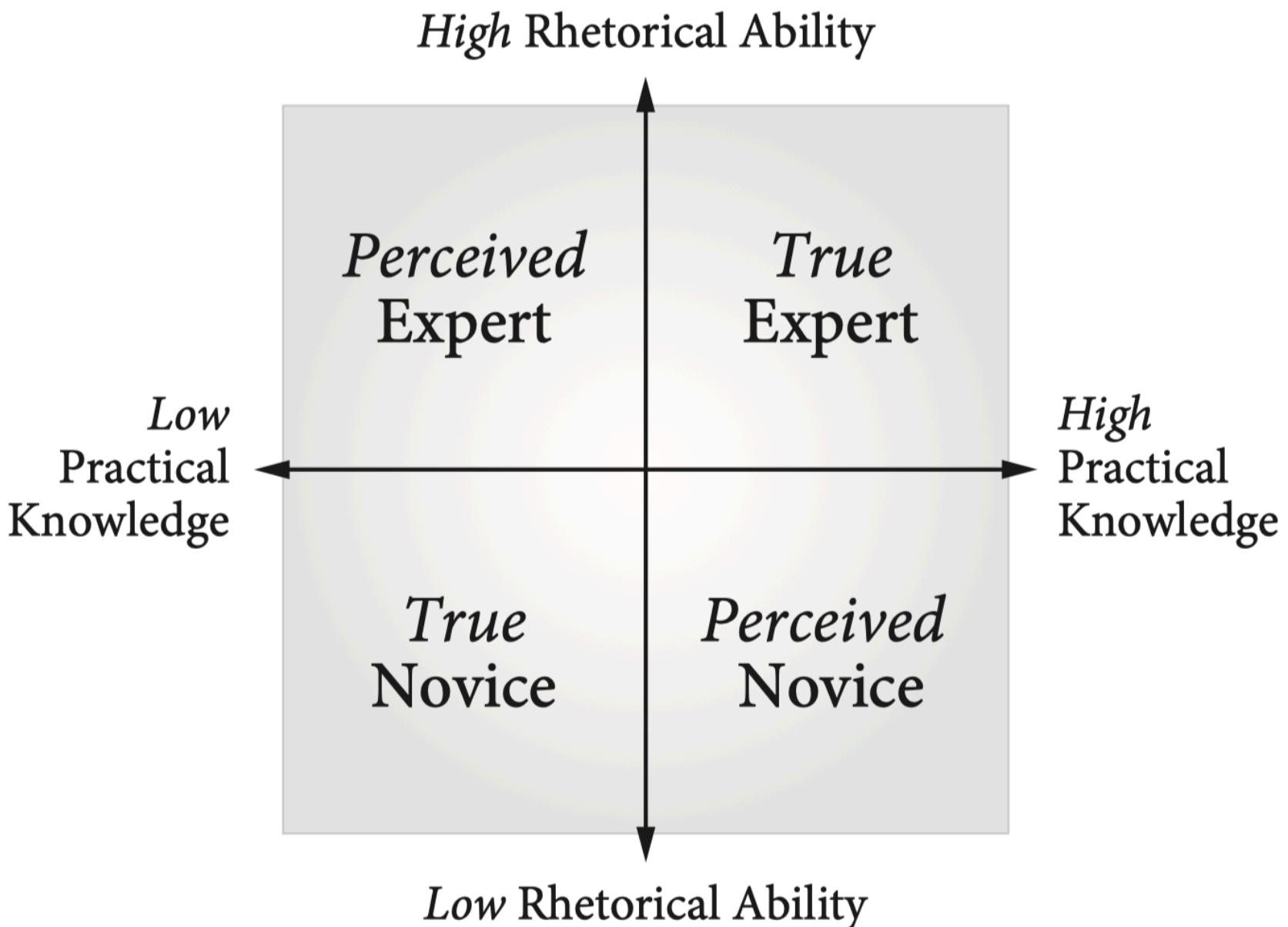
[7] Danescu-Niculescu-Mizil et al., 2013, A computational approach to politeness with application to social factors

Expressing Expertise[8]

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.

High Rhetorical Ability

Perceived Expert	*True* Expert
True Novice	*Perceived* Novice

Low Practical Knowledge ← → *High* Practical Knowledge

Low Rhetorical Ability

[8] Andrist et al., 2013, Rhetorical robots

Expertise Cues[8]	Examples
Goodwill *Wanting the best for the listener.*	**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."
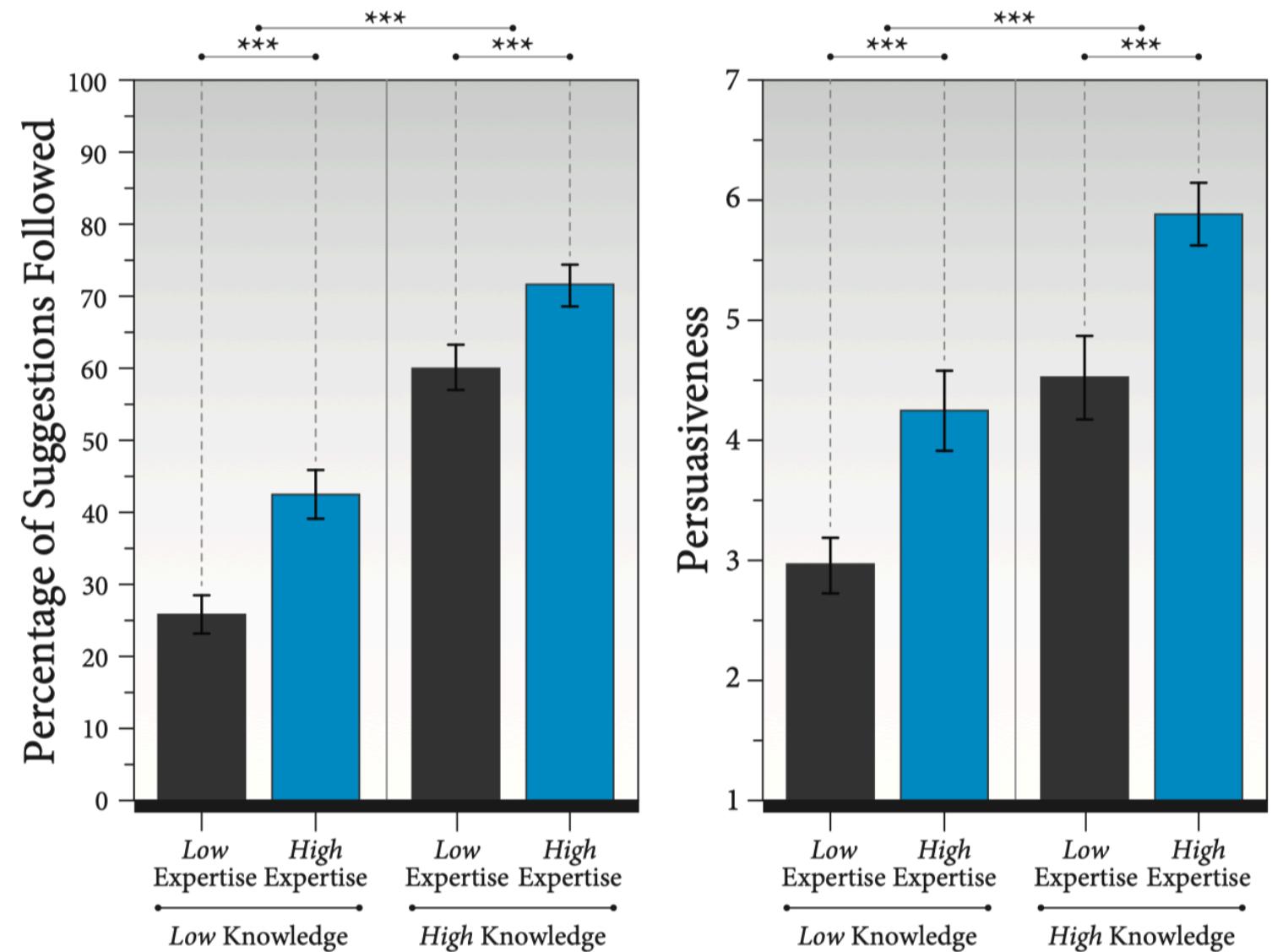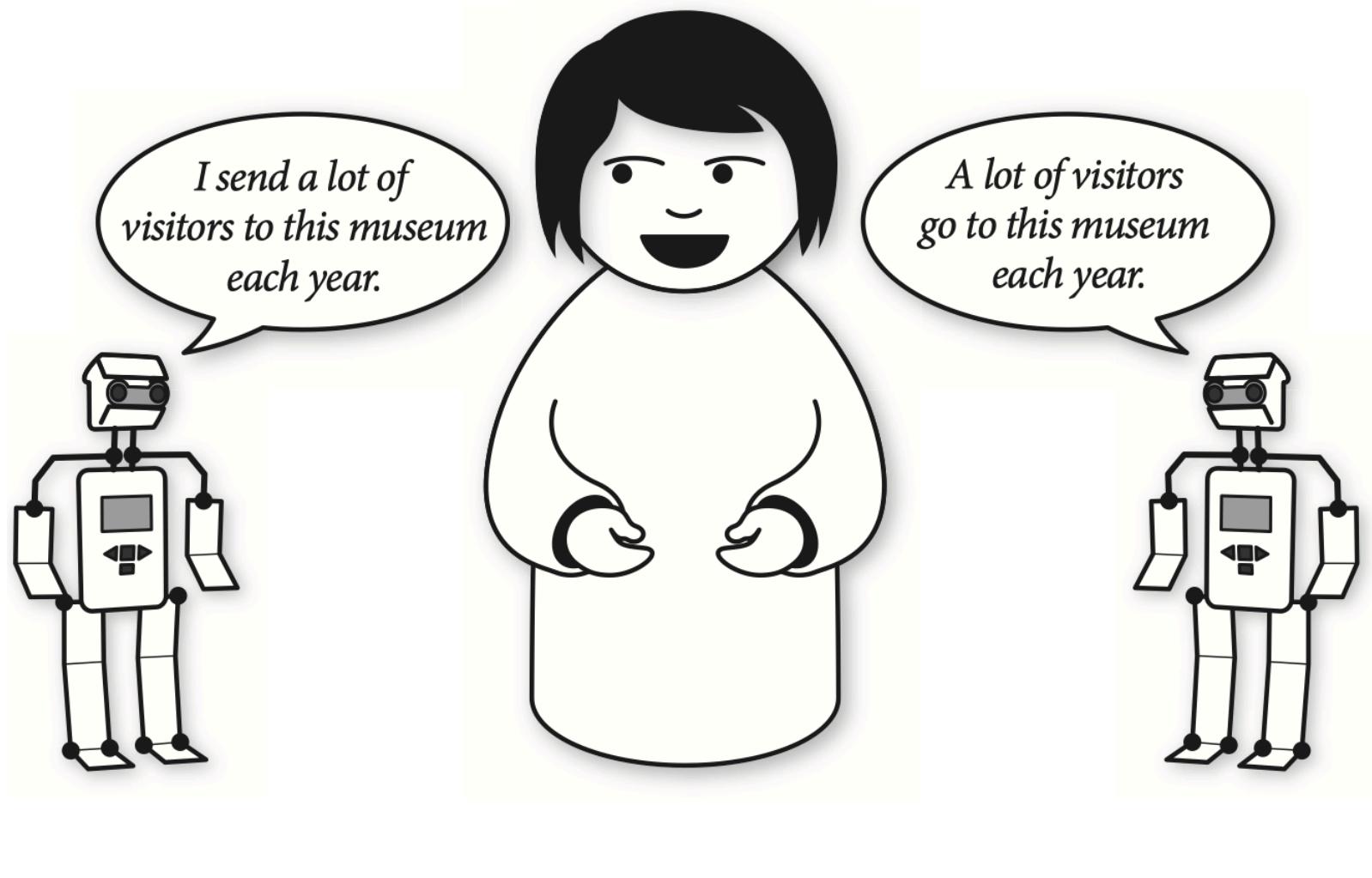
[8] Andrist et al., 2013, Rhetorical robots

Personality in Artificial Agents

What is personality?

Definition: *Personality* refers to individual differences in characteristic patterns of thinking, feeling, and behaving.[9]

There are two prevailing models of personality:
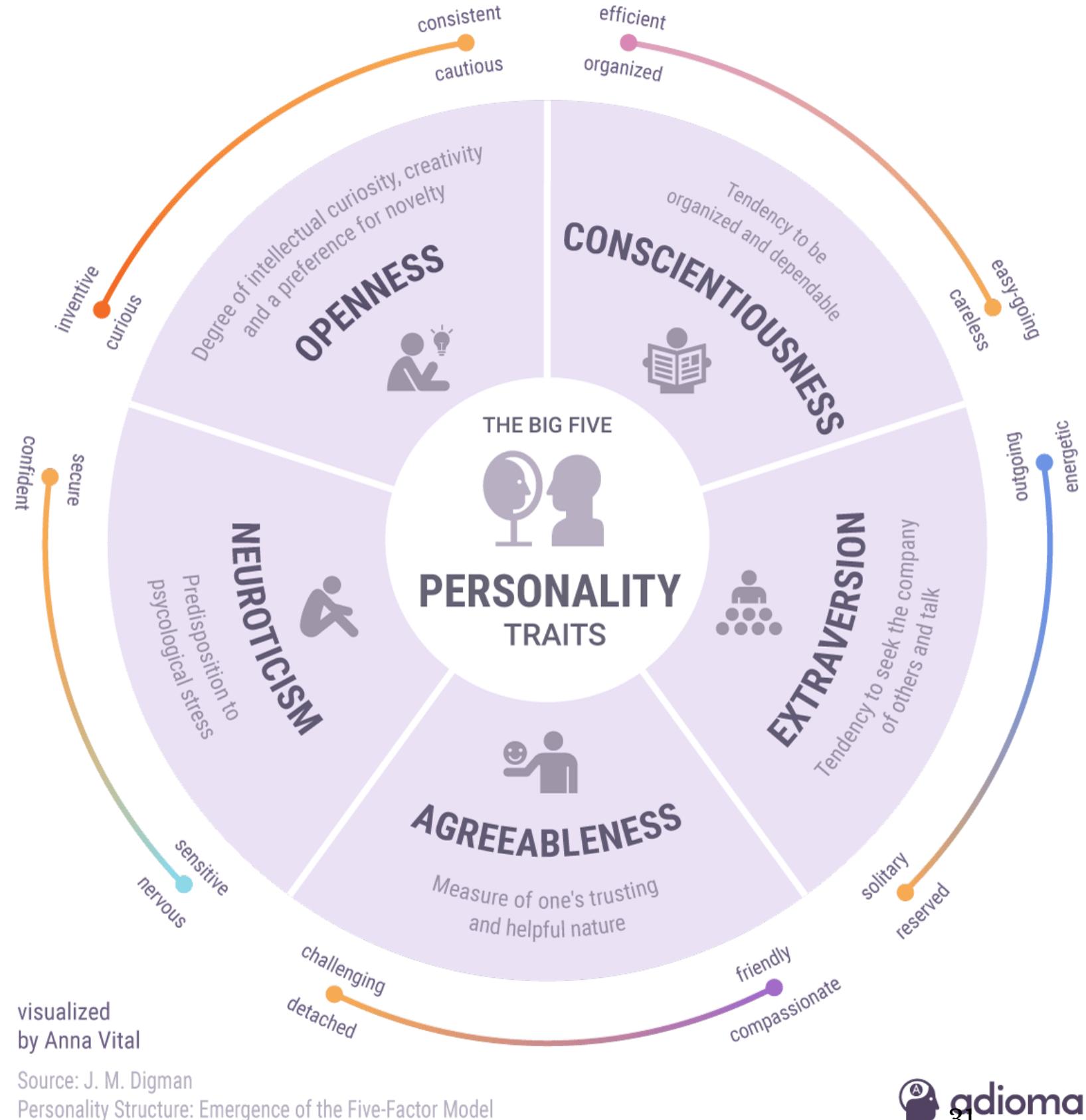
1. Personality traits
2. Personality types

[9] APA

Personality Traits[10] [11]

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

1. Openness
2. Conscientiousness
3. Extraversion
4. Agreeableness
5. Neuroticism
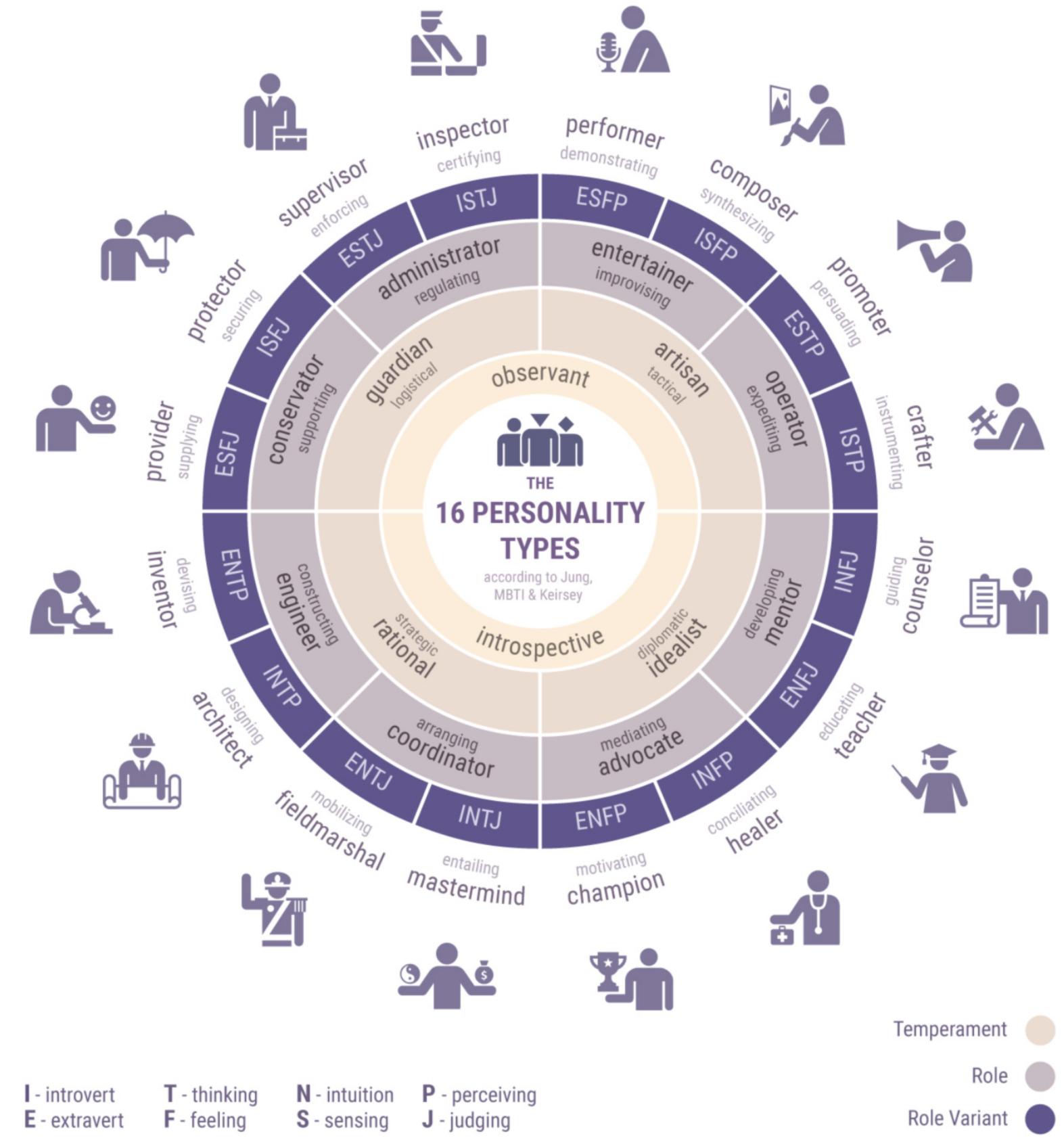
[10] Image source

[11] Live Science

Personality Types[12]

The Myers-Briggs theory, the most commonly accepted set of personality types, posits that there are 16 distinct personalities that vary across four dimensions:
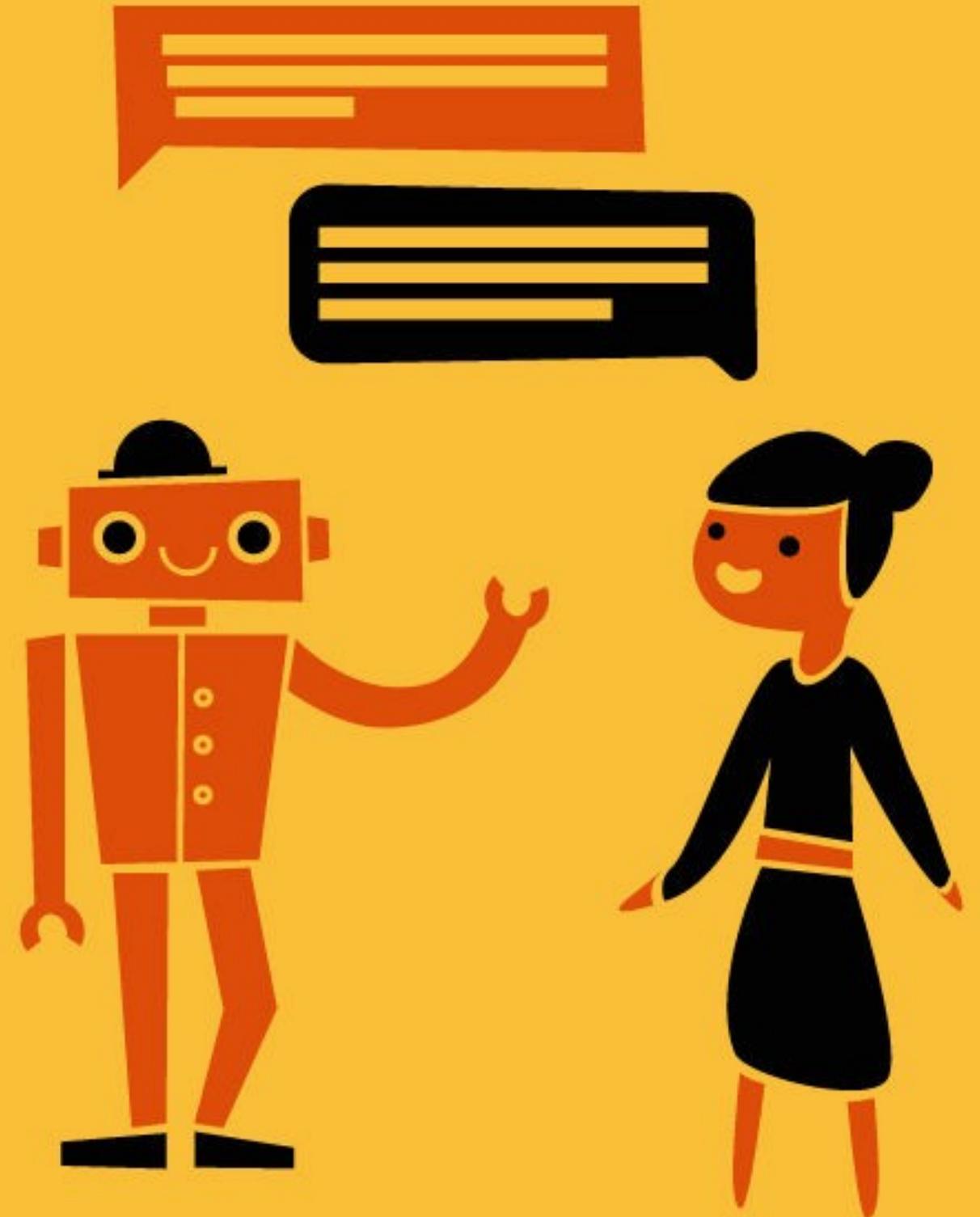
1. Extraversion and introversion
2. Sensing and intuition
3. Thinking and feeling
4. Judgment and perception

Personality Design Strategies[13]

— Personality matching

— Personality expression

— Persona development

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.

Personality Expression

— *Language use*, e.g., politeness cues, expert speech

— *Visible cues*, e.g., gaze, gesture, facial expressions, appearance

— *Marketing*, e.g., target audience, branding

— *Specialization*, e.g., domain knowledge, expertise

Persona Development[14][15]

Definition: Personas are fictional characters, which you create based upon your research in order to represent the different user types that might use your service, product, site, or brand in a similar way.

[14] Image source

[15] Interaction Design Foundation: Personas

Agents and characters can be matched with users at the persona level (as opposed to low-level characteristics or personality).[16]

Right: example use of persona in chatbot behavior

Personalized Bot Flows

Mike — Lives in NYC / Buys: Dress Shirts & Jeans → Polo Classic Dress Shirt

Sofia — Lives in DC / Buys: Shoes & Dresses → Back Dress Collection

Alex — Lives in Miami / Buys:'Big Pony' Polos → Big Pony Polo 25% Off

Jen — Lives in SF / Buys: Home Decor → Blue Designer Sofa
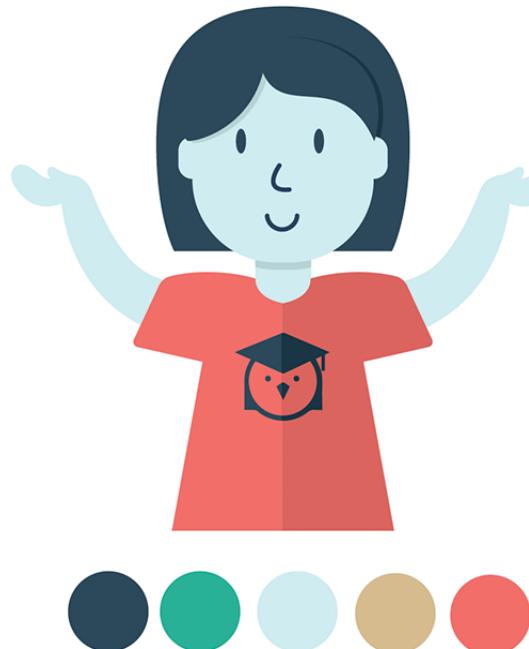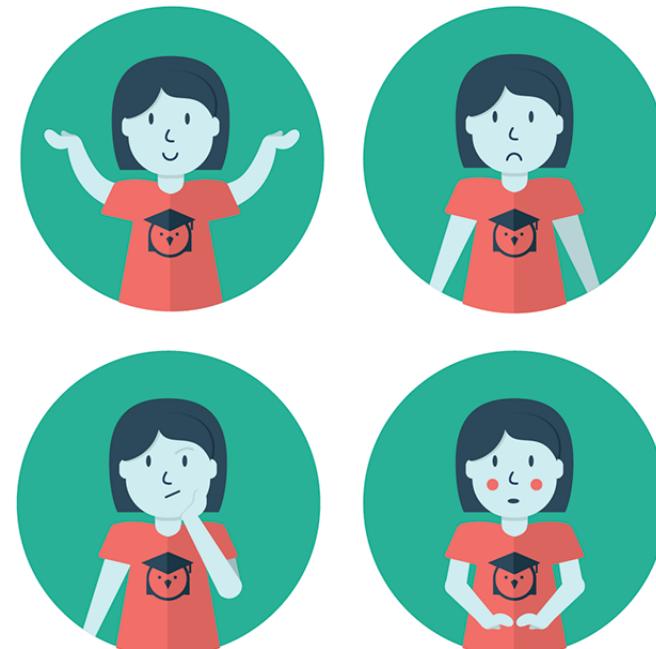
[16] Image source

Linn | Linux Character Bot Illustration

Linn Main Illustration & Color Palette:
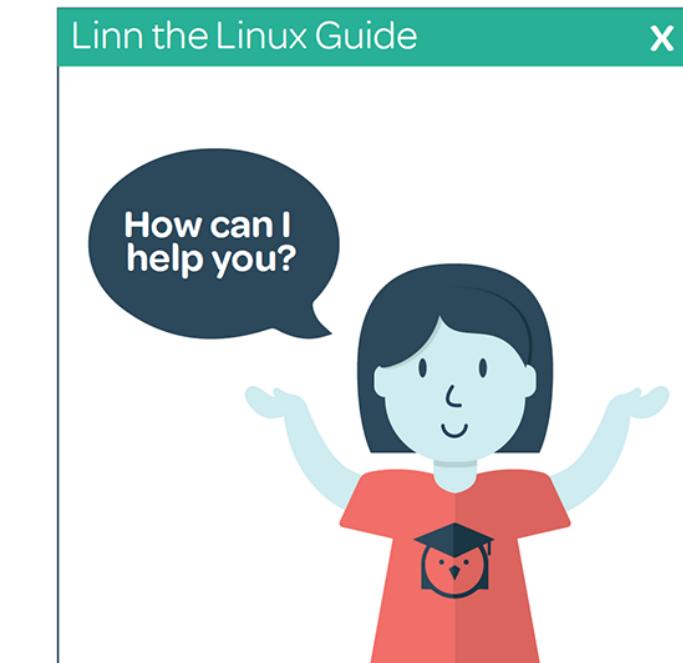
Linn Expressions:
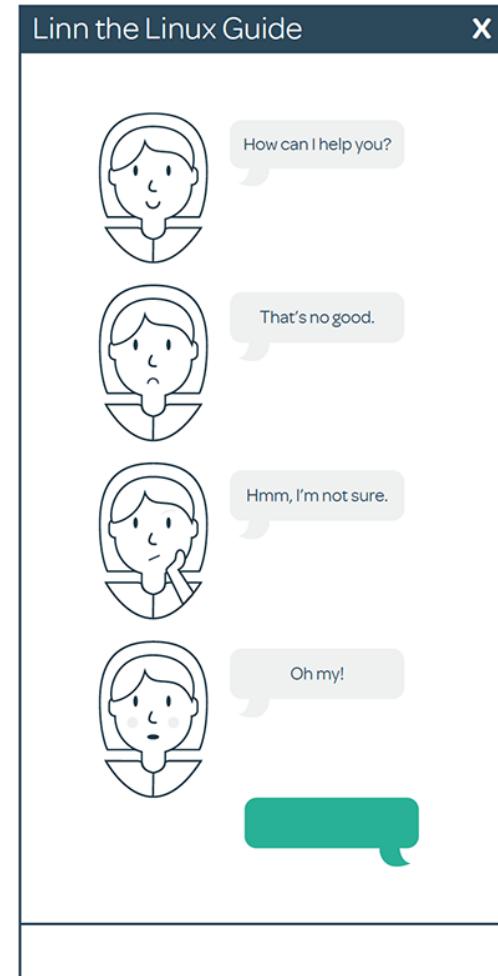
Linn | Chat Examples

Linn Detailed Chat Examples:

Linn the Linux Guide ✕

How can I help you?

Linn the Linux Guide ✕

How can I help you?

That's no good.

Hmm, I'm not sure.

Oh my!

2

4

Further Reading[18]

— The conversation designer's handbook

— A guide to developing bot personalities

— What are & how to create personas

— A closer look at personas

[18] Image source

What did we learn today?

— Computers as Social Actors

— Designing Character Speech

— Personality in Artificial Agents