

Speed Dating with Voice User

Interfaces:

Understanding How

Families Interact and Perceive Voice

User Interfaces in a Group Setting

Jibo

Amazon Echo

Google Home

“Hey Jibo”

“Hey Alexa”

“Hey Google”

Jibo

Amazon Echo

Google Home

“Hey Jibo”

“Hey Amazon”

“Hey Google”

Jibo

“Hey Jibo”

Amazon Echo Spot

“Hey Alexa”

Amazon Echo Show

“Hey Computer”

**social embodiment =
physical embodiment +
interpersonal cues +
social profile (name)**



	Google Home	Amazon Echo	Amazon Echo Show	Amazon Echo Spot	Jibo
Full Name	Google Home	Amazon Echo	Amazon Echo Show	Amazon Echo Spot	Jibo
Name in Study	Google	Alexa (Study 1) Amazon (Study 2)	Computer	Alexa	Jibo
Study	Study 1 & 2	Study 1 & 2	Study 3	Study 3	Study 1, 2 & 3
Wake Word	"Okay Google" "Hey Google"	"Hey Alexa" (Study 1) "Hey Amazon" (Study 2)	"Hey Computer"	"Hey Alexa"	"Hey Jibo"
Physical Form	Cylinder	Cylinder	Rectangle	Circular head+Flag	Circular head+Body
Camera			✓	✓	✓
Mic	✓	✓	✓	✓	✓
Screen			✓	✓	✓
Motion Mechanism	No motion	No motion	No motion	Flag that rotates when hears wake word	Body shifts orientation toward sound + face

FIGURE 1 | Spectrum of VUIs that represents five commercially available VUIs mapped on to varying levels of embodiment going from less socially embodied to more socially embodied.

**how robots influence group
interactions**

**perceptions of the robots and
member satisfaction**

24

information entertainment interpersonal tasks

+ free play

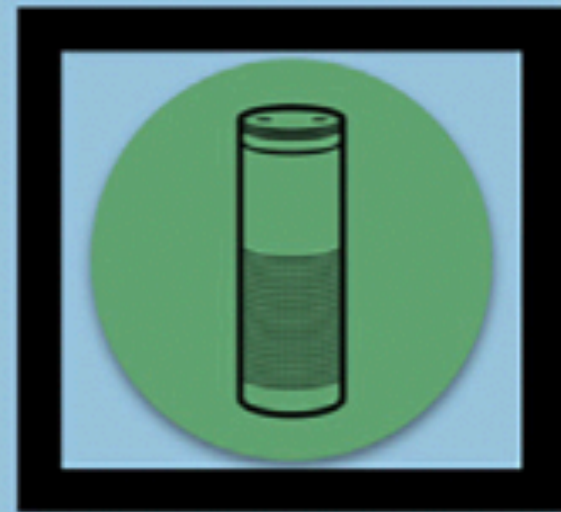
A

 weather	 news	 who was George Washington? where was he born?	 random fact	 time in London	 what is water?
 favorite animal	 agent's name	 age	 favorite food	 favorite movie	 favorite holiday
 dance	 tell me a joke	 self-destruct	 music	 sing me a song	 tell me a poem
 what is the meaning of life?	 go to sleep	 picture	 tell me a secret	 what are you thinking?	 what do you think of the other agents?

LEAST



MOST



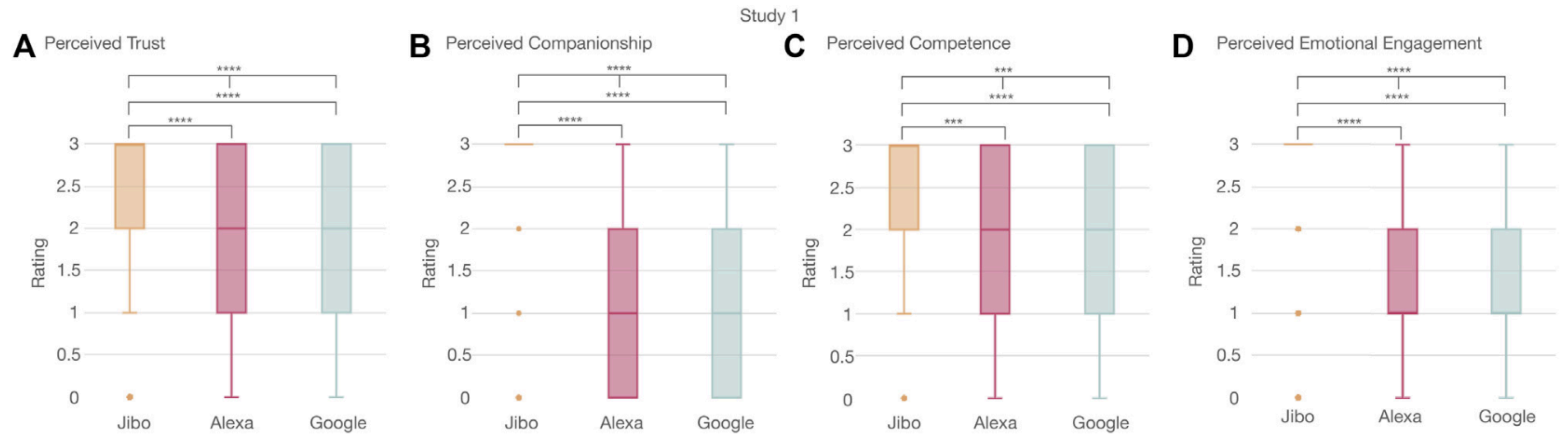


FIGURE 6 | Significant trend in perceived trust, companionship, competence, and emotional engagement of voice agents between participants showing Jibo was significantly perceived as **(A)** more trustworthy, **(B)** more like a companion, **(C)** more competent, and **(D)** more emotionally engaging than Alexa or Google.

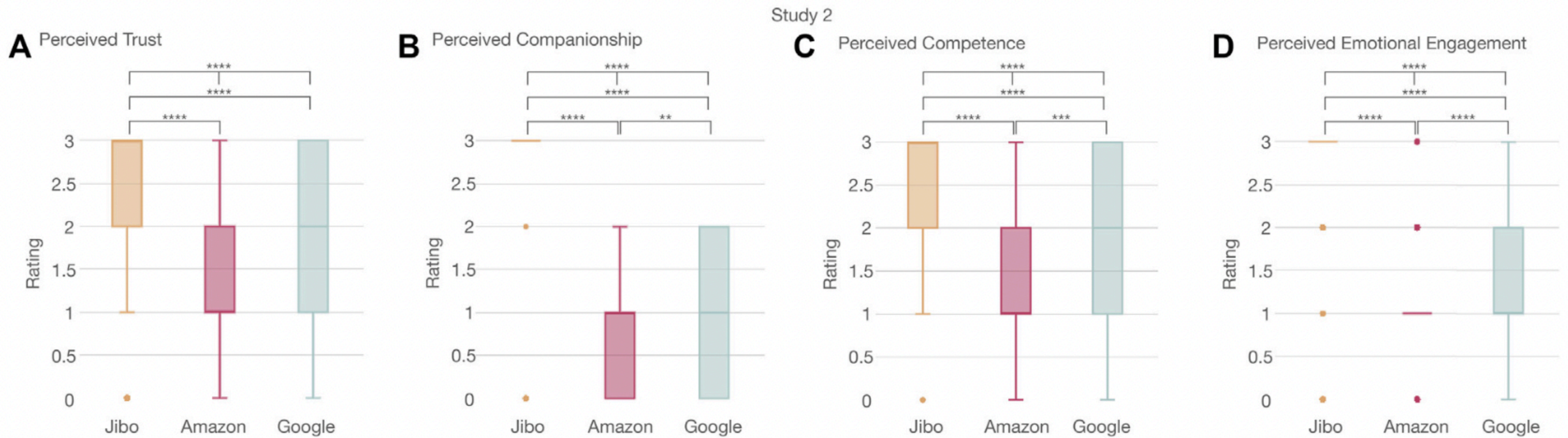


FIGURE 9 | Significant trend in perceived trust, companionship, competence, and emotional engagement of voice agents between participants showing Jibo was significantly perceived as **(A)** more trustworthy, **(B)** more like a companion, **(C)** more competent, and **(D)** more emotionally engaging than Amazon or Google. Google was perceived as **(B)** more like a companion, **(C)** more competent, and **(D)** more emotional engaging than Amazon.

Study 3

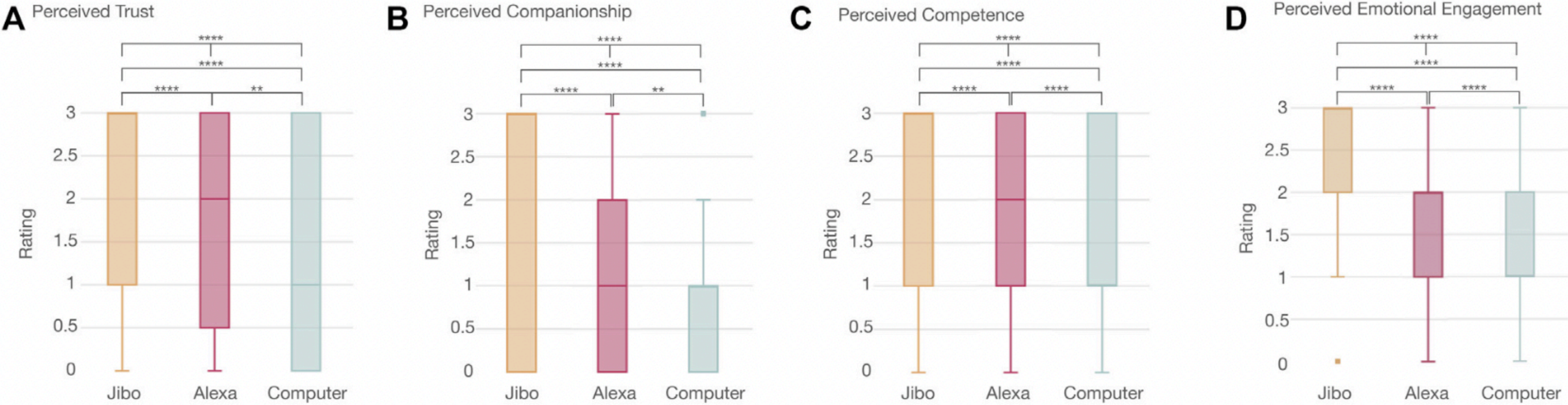


FIGURE 12 | Significant trend in perceived trust, companionship, competence, and emotional engagement of voice agents between participants showing Jibo was significantly perceived as **(A)** more trustworthy, **(B)** more like a companion, **(C)** more competent, and **(D)** more emotionally engaging than Alexa or Computer. Alexa was perceived as **(A)** more trustworthy, **(B)** more like a companion, **(C)** more competent, and (d) more emotionally engaging than Computer.

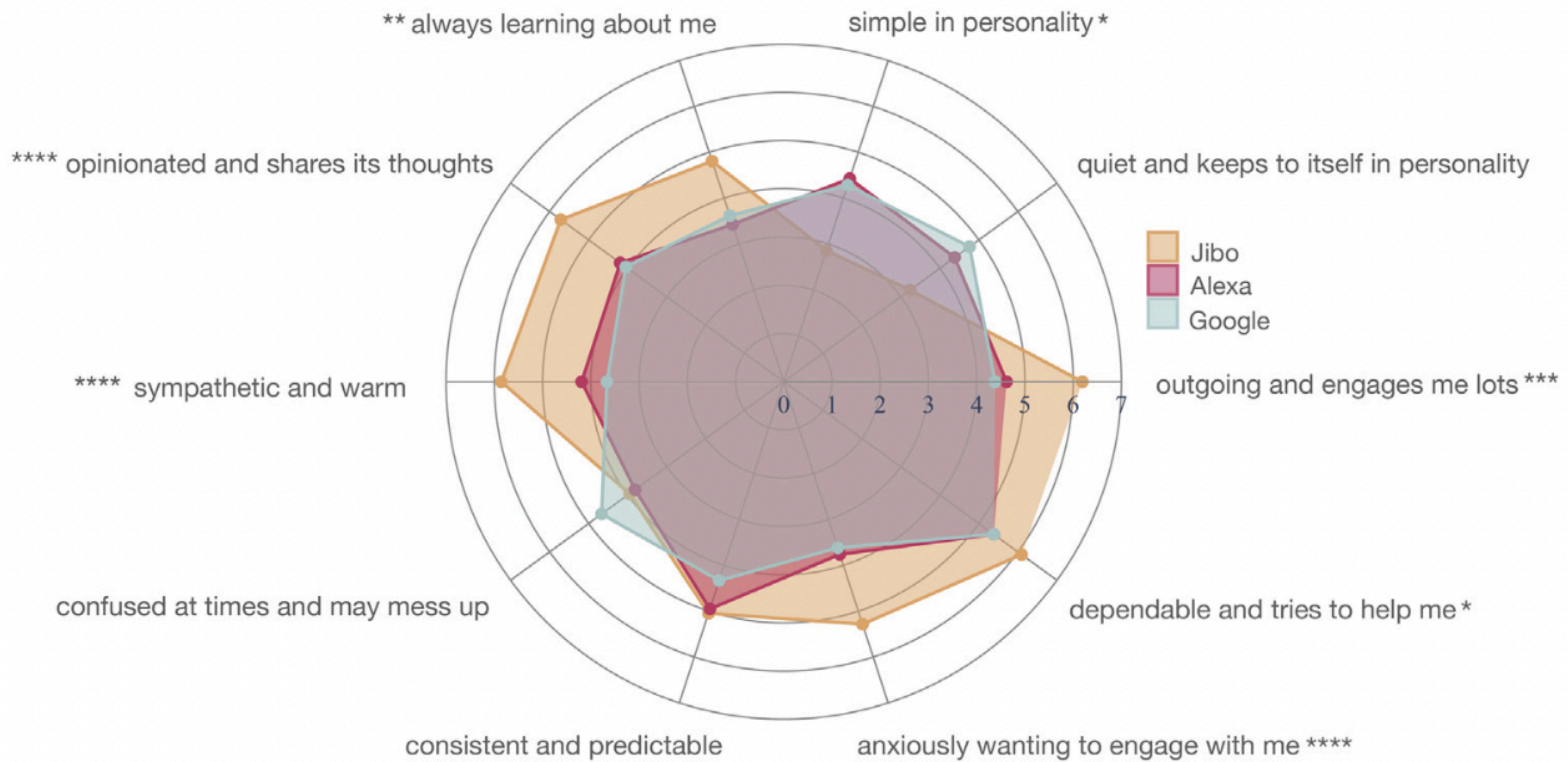


FIGURE 5 | Personality mapping for Jibo (orange), Alexa (red), and Google (blue) in Study 1

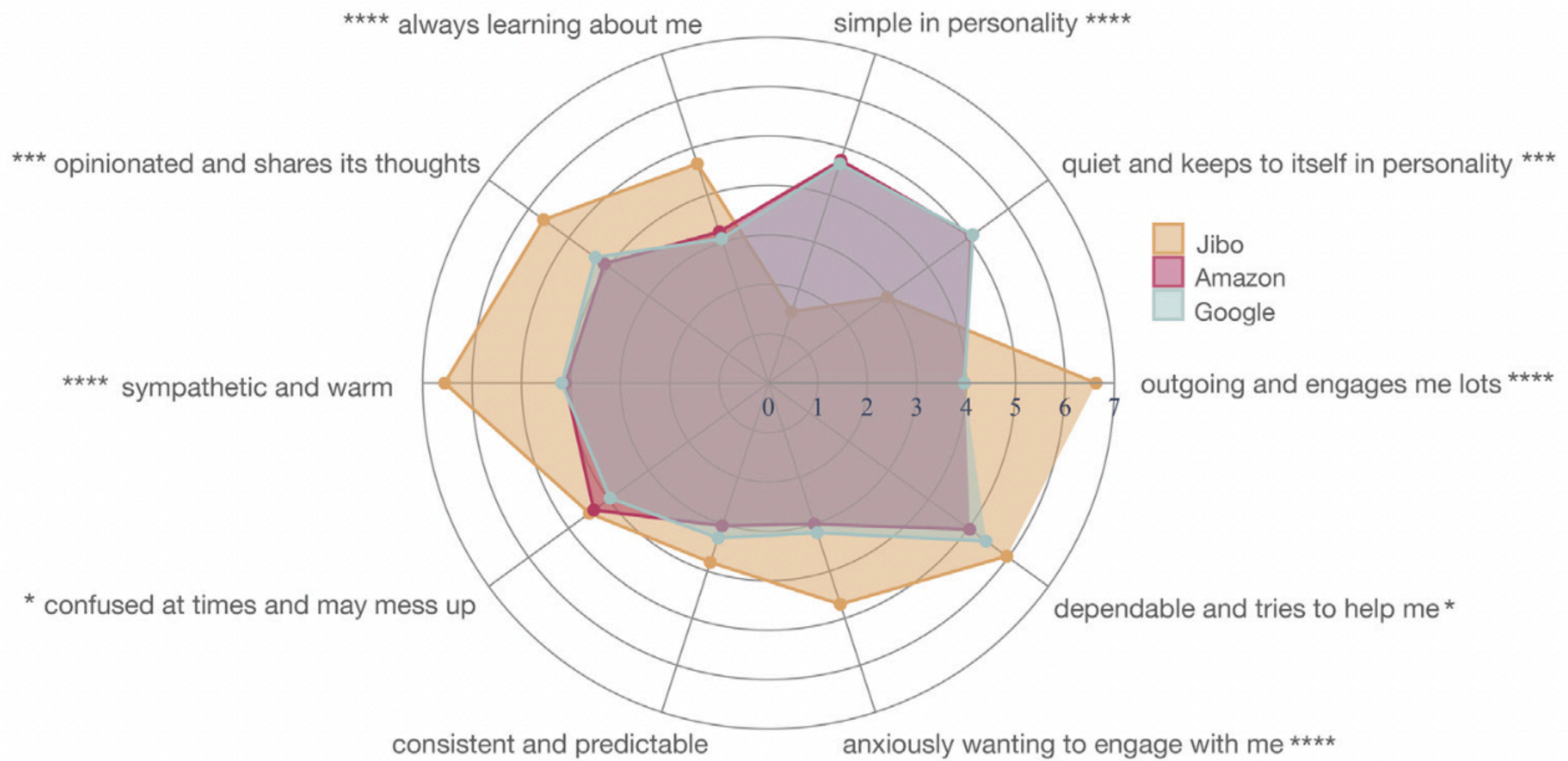


FIGURE 8 | Personality mapping for Jibo (orange), Amazon (red), and Google (blue) in Study 2

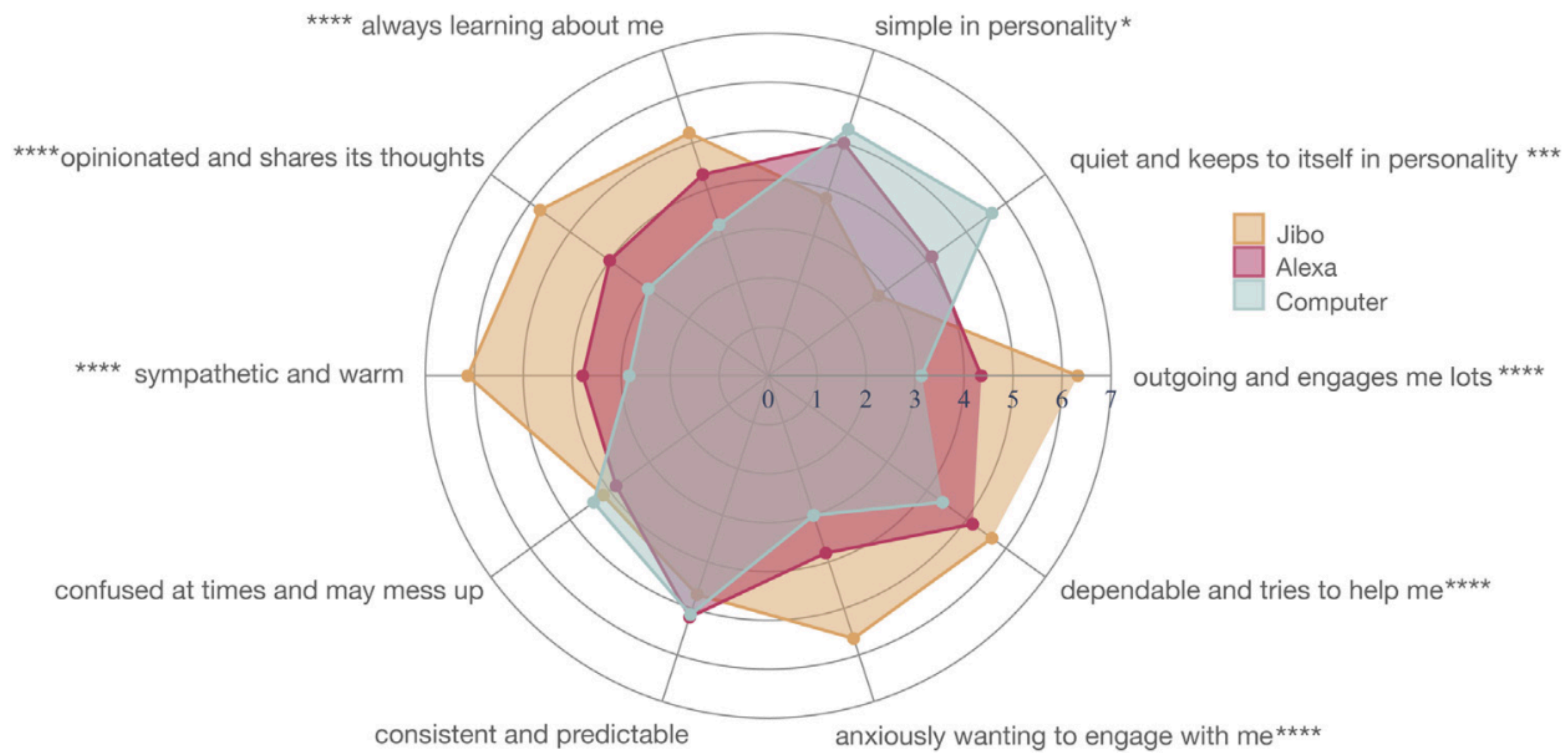


FIGURE 11 | Personality mapping for Jibo (orange), Alexa (red), and Computer (blue) in Study 3

TABLE 2 | Comparing group behaviors observed around the three VUI agents that vary in the degree of social embodiment. The overall result across the combined behaviors and subsequent three behaviors showed statistically significant difference between the agents.

Group behavior	Behavior description	Jibo mean±std	Alexa mean±std	Computer mean±std	Friedman test	Post-hoc wilcoxon with holm correction
Total	Total number of occurrences of coded group behaviors	7.72 ± 6.34	2.91 ± 2.12	2.75 ± 2.91	$\chi^2(2, N = 428) = 27.28$ $p < 1e-04^{****}$	Jibo vs. Alexa: $Z = 4.35, p < 0.01^{****}$ Jibo vs. Comp: $Z = 4.44, p < 0.001^{***}$
Defending an agent	A group member defended an agent's action or response to another group member	1.50 ± 1.00	0.0 ± 0.0	0.0 ± 0.0	$\chi^2(2, N = 4) = 8.0$ $p < 0.05^*$	N/A due to small number of samples
Private conversations	Group member turning to another group member to whisper and/or have a private conversation about agent and/or agent's response	2.7 ± 2.1	1.09 ± 1.12	0.87 ± 1.14	$\chi^2(2, N = 23) = 14.80$ $p < 0.001^{***}$	Jibo vs. Alexa: $Z = 3.13, p < 0.01^{**}$ Jibo vs. Comp: $Z = 3.49, p < 0.001^{***}$
Glancing at others	Group member glances at another group member	5.77 ± 4.74	2.19 ± 1.58	2.19 ± 2.46	$\chi^2(2, N = 31) = 22.94$ $p < 1e-04^{****}$	Jibo vs. Alexa: $Z = 4.20, p < 1e-04^{****}$ Jibo vs. Comp: $Z = 4.06, p < 1e-04^{****}$

TABLE 1 | Comparing reciprocal behaviors observed around the three VUI agents that vary in the degree of social embodiment. The overall result across the combined 37 coded behaviors and the seven behaviors that showed statistically significant difference between the agents are highlighted.

Individual behavior	Behavior description	Jibo mean \pm std	Alexa mean \pm std	Computer mean \pm std	Friedman test	Post-hoc wilcoxon with holm correction
Total	Total number of occurrences of 37 coded behaviors	28.79 \pm 18.89	12.22 \pm 7.96	11.21 \pm 6.52	$\chi^2(2, N = 1711) = 19.22$ $p < 1e-04^{****}$	Jibo vs. Alexa: $Z = 4.75$, $p < 0.01^{**}$ Jibo vs. Comp: $Z = 4.91$, $p < 0.001^{***}$
Agent luring attention	Agent's behavior promotes group member to direct attention and body language to the agent	1.62 \pm 1.19	0.23 \pm 0.44	0.0 \pm 0.0	$\chi^2(2, N = 24) = 19.95$ $p < 1e-04^{****}$	Jibo vs. Alexa: $Z = 2.87$, $p < 0.01^{**}$ Jibo vs. Comp: $Z = 3.06$, $p < 0.01^{**}$
Relevancy	Group member builds on an agent's response	2.1 \pm 1.25	0.55 \pm 0.99	0.75 \pm 0.85	$\chi^2(2, N = 68) = 24.99$ $p < 1e-04^{****}$	Jibo vs. Alexa: $Z = 3.60$, $p < 0.001^{***}$ Jibo vs. Comp: $Z = 3.44$, $p < 0.001^{***}$
Smiling	Group member smiles due to agent's action	4.87 \pm 3.53	1.18 \pm 1.38	1.29 \pm 1.23	$\chi^2(2, N = 241) = 38.66$ $p < 1e-04^{****}$	Jibo vs. Alexa: $Z = 4.77$, $p < 1e-04^{****}$ Jibo vs. Comp: $Z = 4.55$, $p < 1e-04^{****}$
Physical imitation	Group member imitates agent's response using their body	1.25 \pm 0.5	0.0 \pm 0.0	0.0 \pm 0.0	$\chi^2(2, N = 5) = 8.00$ $p < 0.05^*$	N/A due to small number of samples
Looking away	Group member breaks eye-contact with the agent to look elsewhere	2.74 \pm 1.89	1.48 \pm 1.55	1.26 \pm 1.06	$\chi^2(2, N = 148) = 14.02$ $p < 0.001^{***}$	Jibo vs. Alexa: $Z = 2.76$, $p < 0.01^{**}$ Jibo vs. Comp: $Z = 3.33$, $p < 0.001^{***}$
Laughing	Group member laughs at the agent's response	3.86 \pm 3.00	1.21 \pm 2.10	1.0 \pm 1.09	$\chi^2(2, N = 170) = 28.72$ $p < 1e-04^{****}$	Jibo vs. Alexa: $Z = 3.83$, $p < 0.001^{***}$ Jibo vs. Comp: $Z = 4.25$, $p < 1e-04^{****}$
Complimenting	Group member gives a compliment to the agent	1.27 \pm 0.79	0.09 \pm 0.30	0.18 \pm 0.60	$\chi^2(2, N = 17) = 11.64$ $p < 0.01^{**}$	Jibo vs. Alexa: $Z = 2.50$, $p < 0.05^*$ Jibo vs. Comp: $Z = 2.04$, $p < 0.05^*$

**VUIs with higher levels of social
embodiment were
perceived as more trustworthy,
competent, companion-like,
and emotionally engaged**

**Branding, embodiment, and agent
wake word greatly influenced how
lesser socially embodied
agents were interacted with and
perceived**