

Matching Robot Appearance and Behavior to Tasks to Improve Human-Robot Cooperation

Jennifer Goetz
Department of Psychology
University of California,
Berkeley
jgoetz@uclink.berkeley.edu

Sara Kiesler
Human Computer
Interaction Institute
Carnegie Mellon University
kiesler@cs.cmu.edu

Aaron Powers
Human Computer
Interaction Institute
Carnegie Mellon University
apowers@cmu.edu

Abstract

A robot's appearance and behavior provide cues to the robot's abilities and propensities. We hypothesize that an appropriate match between a robot's social cues and its task will improve people's acceptance of and cooperation with the robot. In an experiment, people systematically preferred robots for jobs when the robot's humanlikeness matched the sociability required in those jobs. In two other experiments, people complied more with a robot whose demeanor matched the seriousness of the task.

1. Introduction

We are entering an era in which personal service robots will interact directly with people. Interactive service robots must meet social as well as instrumental goals. They must create a comfortable experience for people, provide appropriate feedback to users, and gain their cooperation. To this end, researchers are studying human-robot interaction in social settings such as homes [12], museums [25], and hospitals [24]. An important question in this regard is how variations in the appearance and social behavior of a robot affect people's responses to the robot. Is the book judged by its cover? We argue "yes." A robot's appearance and behavior provide cues that influence perceptions of the robot's propensities, and assumptions about its capabilities. The present controlled experiments examined how people perceive and interact with humanoid service robots whose appearance and demeanor we varied systematically.

Theoretical Background

An extensive believable agent literature [e.g. 3, 17, 22] addresses people's interaction with embodied agents presented on a computer display. This work suggests that robotic assistants, to be effective, should exhibit naturalistic behavior and appropriate emotions, and should require little or no learning or effort on the part of the user. In robotics, this premise has stimulated technological advances in biologically-inspired intelligent robots [2]. Many advances have been made in producing robots whose behavior exhibits recognizable emotions such as surprise and delight [4]. Research by Nass and his

colleagues suggests that a computer's demeanor should follow social rules of human-human interaction [18]. Furthermore, the demeanor of the computer is likely to elicit social responses in the user [21]. Studies examining humanlike computer agents support these arguments [19] and suggest that they would be generalizable to robots.

Psychological research suggests that people's initial responses to a robot will be fast, automatic (unconscious), and heavily stimulus- or cue-driven [1]. Even as infants, people automatically perceive objects that make lifelike movements as living things [23]. We argue that humanoid robots convey animistic and anthropomorphic cues that evoke automatic perceptions of lifelikeness in the robot. These perceptions will lead to people making attributions of ability and personality to the robot. In turn, their social responses and expectations will be shaped by these initial attributions. Hence, the nature of a humanoid robot's appearance and demeanor should mediate people's acceptance and responses to them.

2. Studies of Acceptance and Compliance

As noted above, personal service robots must interact with those they serve, and will need to elicit acceptance and compliance from them. The considerable social psychological literature on compliance suggests some directions for design. First, we know that people respond positively to attractive and extraverted people [6] and to a happy, enthusiastic approach [8]. This work suggests a *positivity hypothesis* – the more attractive a robot looks and the more extraverted and cheerful its behavior is, the more people will accept and comply with the robot.

In contrast to the positivity hypothesis is the *matching hypothesis* – that the appearance and social behavior of a robot should match the seriousness of the task and situation. Imagine, for example, a future service robot that delivers bedside medications in a hospital. Studies in medical settings suggest that good cheer and enthusiasm do not always work well. In one study, less consultative and accepting, more authoritative physicians were more effective in gaining patients' confidence [16]. Nurses were more effective when they matched their behavior to the patient's situation [20]. Humor by medical residents was found to be effective only when appropriately matched to

less serious situations [11]. Finally, physicians who expressed anger or deep concern were more effective in obtaining patients' compliance with important treatments than those who acted more lightly [13]. This work supports the matching hypothesis.

Study 1: Preferences for humanlike robots in jobs

The matching hypothesis suggests that a more humanlike appearance is a better match for jobs that are more, rather than less, social in nature. We predicted that people would prefer a humanlike robot in social jobs such as a dance instructor and a more machinelike robot in less social jobs such as a night security guard. We tested this hypothesis in an experiment.

Method

We created 12 2D robotic heads with three levels of humanlikeness – human, “midstage,” and machine (a follow up study with more information on what makes robot faces humanlike is in [9]). To provide a mixture of styles, we also varied whether these robots were more adult looking or more youthful looking [27] but did not formulate hypotheses for these differences, and do not discuss them further. We also created more feminine and more masculine heads of each type, but these were kept constant for each participant. Half of the participants only judged feminine robots and half judged only masculine robots. A pilot study confirmed these manipulations were effective.

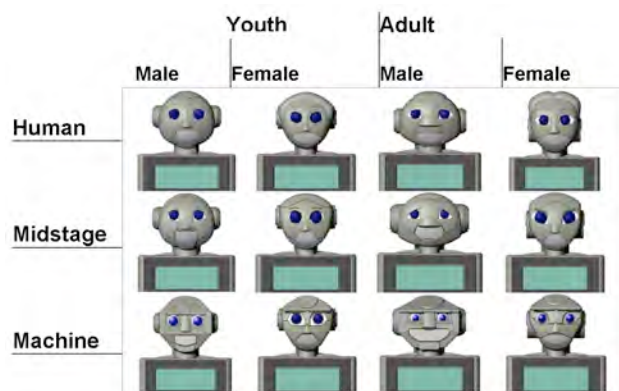


Figure 1. Robots in survey.

Participants were 108 college and graduate students. Their average age was 26 (SD = 8 years); 60% were male. In an online survey, these participants made a series of choices between two of the robots at a time (both feminine or both masculine). Participants were asked which robot would be suitable for service robot jobs chosen from the Strong-Campbell Interest Inventory [7], which classifies jobs based on the interests of people who do them. The analyses of the data were performed separately for the job groups and robot gender. We used mixed models repeated measures ANOVAs. The dependent variable was the number of times a particular robot was selected. Due to

the large number of effects tested, only results for the main effect of style (humanlikeness vs. machinelikeness) and the hypothesized interaction of style and Strong category are reported here.

Results

For the female robots, both the main effect of style and the interaction of style and Strong category were significant ($p < .0001$). Overall, participants preferred the humanlike robots to the machinelike robots for most jobs, including the following jobs and Strong categories: actress and drawing instructor (Artistic), retail clerk and sales representative (Enterprise), office clerk and hospital message and food carrier (Conventional), aerobics instructor and museum tour guide (Social). However, as hypothesized, participants preferred the machinelike robots over the humanlike robots for jobs including lab assistant and customs inspector (Investigative) and for soldier and security guard (Realistic).

Patterns for the masculine looking robots were not as strong but were generally in the same direction. Participants slightly preferred humanlike robots for Artistic and Social job types, but they preferred machinelike robots for Realistic and Conventional job types.

These results support the matching hypothesis, in that humanlike robots were more preferred for jobs that require more social skills (when these jobs are performed by people, according to the Strong Interest Inventory). Our results also imply that effects of robotic appearance are not only systematic, but might be predicted from population stereotypes.

Study 2: Compliance with a playful or serious robot

We conducted this study to explore whether a service robot's social demeanor would change people's compliance with the robot's requests. The positivity hypothesis predicts that a cheerful, playful robot will elicit more compliance in users, whereas the matching hypothesis predicts more subtle effects – that a cheerful, playful robot will elicit more compliance if the task context is related to entertainment or fun, but that a more serious or authoritative robot will elicit more compliance if the task context is more serious, urgent, or disagreeable such as getting a chore done, taking medication or sticking to an exercise routine.

Because physical exercise is a task that is good for people but most fail to do regularly [5, 14] we prototyped a robot's social behavior to create two types of demeanor – playful versus serious. The robot asked participants to perform some exercise routines with one or the other of these demeanors. If the positivity hypothesis is correct, then participants who interact with the playful robot should comply more with the robot's exercise requests than will those who interact with the serious robot. If the matching hypothesis is correct, then participants who

interact with the serious robot should comply more with the robot's exercise requests.

Method

The procedure involved a Wizard of Oz interaction between a participant and a humanoid robot. After obtaining informed consent, the experimenter left the participant alone with the robot. The robot initiated a brief social conversation with the participant and instructed the participant in a few exercises. Then the robot asked the participant to make up his or her own exercise routine and perform it. The independent variable was the robot's demeanor during these interactions, as manipulated through its speech to be playful or serious, but with the same content. The dependent variable was the amount of time that participants exercised by themselves when the robot asked them to create and perform an exercise routine.

Twenty-one participants were randomly assigned to interact with either the serious ($n = 11$) or the playful ($n = 10$) robot. Participants averaged 25 years old. There were 9 females and 13 males. (We report only results from native English speakers because nonnative speakers frequently failed to understand the robot's simulated speech.)



Figure 2. Robot and participant.

Robot

The robot used in the study was the Nursebot robot, Pearl (www.cs.cmu.edu/~nursebot/). Figure 2 shows the robot and participant, as the experimenter leaves them alone to interact. The robot used speech to interact with the participant but did not move about.

Procedure

After a participant arrived, the experimenter left the participant with the robot, entering an adjacent room. The experimenter could hear the interaction between the participant and the robot through a microphone on the robot but could not see the interaction. The interaction was videotaped with two cameras – one in the eye of the robot

and one placed in the room. The robot gave instructions for the experiment, following either the playful or serious script. For experimental control, the scripts were designed to stay the same no matter how a participant answered. Excerpts from the two scripts follow.

Playful Script

Playful Robot: Do you like to exercise?

Participant: [answers]

Playful Robot: That's ok. These are fun--you'll love them. Let's start. I want you to breathe to warm up. Do you know how to breathe?

Participant: [answers]

Playful Robot: Ha ha ha! I hope so. Ready to start?

Participant: [answers]

Playful Robot: Close your eyes. [wait] Relax. [wait] Breathe in.[wait]. Don't forget to breathe out. I don't want you to pass out!

Serious Script

Serious Robot: Do you exercise?

Participant: [answers]

Serious Robot: It is very important to your health. I would like to have you do some exercises now. Would that be okay?

Participant: [answers]

Serious Robot: Good, try to do everything that I say as best you can. Let's start with a breathing exercise. Are you ready?

Participant: [answers]

Serious Robot: Close your eyes. [wait] Relax. [wait] Breathe in. [wait] Breathe out. [wait] Are you feeling relaxed?

An experimenter's assistant initiated and controlled the timing of the script in the control room, using in-house software that interfaces with the Festival Speech Synthesis System. Both scripts asked participants to close their eyes and breathe, stand up, stretch, and touch their toes. The experimenter then entered the room temporarily to ask the participant to complete a questionnaire about the robot and its personality [15]. Next, the compliance request began. The robot asked the participant to do more exercises. It asked the participant to stand on one foot and do a series of balancing exercises. Then, the robot asked the participant to make up an exercise routine with stretches. The robot instructed participants to continue as long as they could and gave encouraging remarks (i.e. "Good job") about every 5 seconds. When participants said that they were finished or tired, the robot confirmed they were finished. It then thanked them for their help. The experimenter entered, administered a final questionnaire, and thanked participants.

Results

In the voluntary exercise portion of the interaction, participants made up a routine and exercised for the robot

an average of 40 seconds. The distribution of exercise time was positively skewed, so we performed a natural log transformation of the data. According to the analysis of variance, participants exercised longer when the robot was serious than when the robot was playful, supporting the matching hypothesis (means = 53 vs. 25 seconds, respectively, $p=.01$).

The results of the questionnaire indicate how participants perceived the robot before the robot asked them to exercise on their own. Table 1 (Study 2) shows that participants rated the serious robot as significantly higher than the playful robot in conscientiousness (a Big Five trait [15]), and also rated it as smarter but less playful and less witty than the playful robot. They also rated the playful robot as slightly more obnoxious.

Discussion

Although the results of Study 2 were consistent with the matching hypothesis, we had to perform another study to show that people would comply with a playful robot more than a serious robot if they were doing an enjoyable or entertaining task. Study 3 addressed this issue.

Study 3: Compliance with a playful or serious robot on an entertaining vs. serious task

In this study, we tested the matching hypothesis directly by comparing compliance with a robot on two tasks, the exercise task and a jellybean recipe task that required participants to taste different flavors of high-quality jellybeans, and to create “recipes” such as coconut pie and banana nut sundae. We predicted participants would comply more with the playful robot than with the serious robot in the jellybean task condition, and more compliance with the serious robot than with the playful robot in the exercise task condition.

Method

Forty-seven participants were randomly assigned to one of four conditions in a 2 (robot demeanor) x 2 (task context) factorial design. Participants averaged 23 years old. There were 23 females and 24 males. All were native English speakers. Two experimenters conducted the study.

The experimental procedure was generally the same as that used in the first study. We created serious and playful robot scripts for the jellybean task that mirrored the scripts used for the exercise task. In the jellybean condition, participants were given trays of high-quality jellybeans of various flavors. In Phase 1 of the study, the robot asked participants to guess the flavors of differently colored and flavored jellybeans. The participants then completed the first questionnaire. In Phase 2, the robot asked participants to make up combinations of jellybean flavors. The robot led participants through an example and then asked them to make up their own recipes.

The exercise scripts were mostly the same as those used in Study 1 with an exception. To make the two tasks comparable, we changed the exercise instructions to say “Please make up *as many* exercises as you can.” (In Study 2, we asked participants to exercise “as long as you can.”) In the jellybean condition we asked participants “make up *as many* combinations as you can.” These changes in the robot’s instructions increased overall compliance in Study 3 as compared with Study 2.

Results

As shown in Table 1 (Study 3), participants complied with the robot’s request an average of 180 seconds in the jellybean task condition and 110 seconds in the exercise task condition. The distributions were positively skewed, so we performed a natural log transformation of the data.

Table 1. Mean compliance with the robot, and mean perceptions rated on 5 pt scales (standard deviations in parentheses).

Compliance & perceptions	Exercise task – Study 2		Long Exercise task – Study 3		Long Jellybean task – Study 3	
	<i>Robot Demeanor</i> Playful	Serious	<i>Robot Demeanor</i> Playful	Serious	<i>Robot Demeanor</i> Playful	Serious
Compliance secs.	25 (.18)	53 (.29)**	95 (.162)	125 (.167)	218 (.162)	145 (.119)*
Number diff. acts			3.3 (.2.9)	3.7 (.2.3)	8.7 (.6.6)	5.8 (.3.3)
Conscientiousness	3.7 (.6)	4.2 (.4)**	3.5 (.4)	3.8 (.5)*	3.8 (.7)	3.8 (.7)
Extraversion	3.9 (.6)	3.6 (.5)	3.8 (.4)	3.2 (.4)***	3.9 (.4)	3.3 (.6)***
Entertaining	4.4 (.92)	4.2 (.60)	4.4 (.51)	3.7 (1.1)**	4.6 (.7)	4.3 (.6)
Friendly	4.6 (.50)	4.3 (.47)*	4.5 (.52)	3.9 (.67)**	4.5 (.69)	4.1 (.79)
Obnoxious	2.9 (1.51)	1.9 (1.14)*	2.9 (1.44)	2.3 (1.14)	2.4 (1.29)	2.3 (1.37)
Playful	4.2 (.60)	3.1 (.94)***	3.8 (.87)	3.3 (1.06)	3.9 (.94)	3.8 (1.06)
Witty	3.6 (1.12)	2.4 (1.12)**	2.9 (1.24)	2.6 (1.00)	3.6 (1.04)	2.6 (1.08)**
Intellect	3.2 (.48)**	3.6 (.42)**	3.0 (.62)	3.5 (.34)**	3.1 (.44)	3.4 (.72)

* $p<.10$ ** $p<.05$ *** $p<.01$ **** $p<.001$ (playful versus serious comparisons, within each study).

Note: Attributes on which the robots did not differ were agreeableness, neuroticism, open to experience, likeability, annoying, funny, unpleasant, efficient, technological, safe, low maintenance, durability, look human, and act human. All ratings were on were 5 point scales. Significant results are bolded for readability.

An analysis of variance (controlling for experimenter) showed that the participants did the jellybean task longer. This finding supports our premise that the jellybean task was intrinsically more enjoyable than the exercise task. The interaction between script and task was marginally significant in the direction predicted ($F [1, 42] = .10$). That is, the playful robot elicited more compliance than the serious robot did in the jellybean condition, but the serious robot elicited more compliance than the playful robot did in the exercise condition. The playful robot in the jellybean condition elicited the most compliance ($F [1, 42] = 7.6, p < .01$). A similar analysis of variance on the number of unique tasks participants gave results in the same direction (see Table 2). The chart in Figure 3 shows the results of both compliance experiments together.

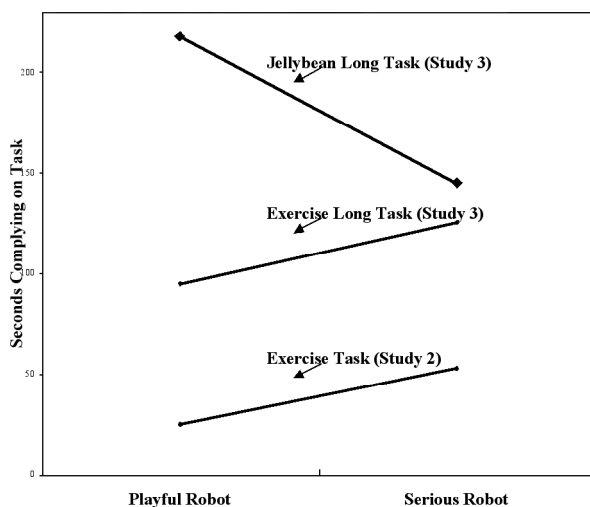


Figure 3. Compliance in Studies 2 and 3.

Participants rated the playful robot as more extraverted on the Big Five personality scale, more playful, more entertaining, friendlier, and wittier than the serious robot. They rated the serious robot higher on the intellect scale, though not significantly so in the jellybean conditions.

3. General Discussion and Conclusion

Several limitations apply to our studies. First, due to the complications of conducting an experiment with a research robot, comparatively few participants were included in each condition of the experiments; hence our statistical power was limited. Second, the results apply only to native English speakers. As we noted earlier, the robot's speech was unclear to nonnative English speakers. Third, these were lab experiments with college and graduate students and perhaps not generalizable to the general public.

Our results do suggest strongly, however, that a robot's appearance or demeanor systematically influences

people's perceptions of a robot, and their willingness to comply with the robot's instructions. These perceptions and responses are evidently elicited by social cues embodied in the robot and are framed by people's expectations of the robot's role in the situation. Hence, participants in our studies did not find the more humanlike, attractive, or playful robot more compelling across the board. Instead, they expected the robot to look and to act appropriately, given the task context. A robot that confirmed their expectations also increased their sense of the robot's compatibility with the robot's job, and their compliance with the robot. Our results imply that the design of a robot's form and interaction behaviors will be an important step in the development of effective personal service robots.

Twenty years ago, Pamela McCorduck urged researchers to create a wonderful "geriatric robot" that would serve as an aide, coach, and good listener, rolled into one "down-home useful" machine [10, pp. 92-93]. Computer technologies are not yet up to that capability, but rapid progress is being made on many fronts ranging from machine learning to materials. Understanding how to design the human-robot interface is an important component of this effort. A key problem in this domain is to find the best mix of machinelike and humanlike interface attributes to support people's goals and a robotic assistant's functionality. This work on a robot's appearance and demeanor represents an early empirical step in a longer agenda.

Acknowledgements

This research was supported by National Science Foundation grants NSF IIS-0121426. Additional thanks to Allyson Pottmeyer, Matt Goodworth, Sebastian Thrun, Nick Roy, Joelle Pineau, Jamie Schulte, Mike Montemerlo, Greg Baltus, Greg Armstrong, Kyle Kirby, Laura Martin, Ronette Killcrece, and Anthony Porcelli.

References

- [1] Aarts H, Dijksterhuis A. Attitudes and social cognition - habits as knowledge structures: Automaticity in goal-directed behavior, in *Journal of Personality and Social Psychology*, vol. 78, 2000, pp. 53-63.
- [2] Bar-Cohen Y, Breazeal C. Biologically-inspired robots, SPIE Press, Bellingham, WA, 2003.
- [3] Bates, J. The role of emotion in believable agents, in *Communications of the ACM*, vol. 37, 1994, pp. 122-125.
- [4] Breazeal, C.L. Designing sociable robots, MIT Press, Cambridge, MA, 2002.
- [5] Bock BC, Marcus BH, Pinto BM, Forsyth LH. Maintenance of physical activity following an

- individualized motivationally tailored intervention, in *Annals of Behavioral Medicine*, vol. 23, 2001, pp. 79-87.
- [6] Bonner BL. The effects of extroversion on influence in ambiguous group tasks, in *Small Groups*, vol. 31, 2000, pp. 225-244.
- [7] Campbell D, Crichton L, Hansen JI, Webber PA. New edition of the SVIB: The Strong-Campbell Interest Inventory, in *Measurement & Evaluation in Guidance*, vol. July 7(2), 1974, pp. 92-95.
- [8] Cialdini RB. *Influence: the psychology of persuasion*, William Morrow, New York, 1993.
- [9] DiSalvo CF, Gemperle F, Forlizzi J, Kiesler S. All robots are not created equal: The design and perception of humanoid robot heads, in *Proceedings of the DIS Conference*, London, England, June 25-28, 2002, pp. 321-326.
- [10] Feigenbaum EA, McCorduck P. *The fifth generation: Artificial intelligence and Japan's computer challenge to the world*, Addison-Wesley, Reading, MA, 1982.
- [11] Francis L, Monahan K, Berger C. A laughing matter? The uses of humor in medical interactions, in *Motivation & Emotion*, vol 23, 1999, pp. 155-174.
- [12] Friedman B, Kahn PH, Hagman, J. Hardware companions? What online AIBO discussions forums reveal about the human-robotic relationship, in *Proceedings of the conference on Human Factors in Computing Systems (CHI 2003)*, vol. 5, 2003, pp. 273-280.
- [13] Hall JA, Roter DL, Rand CS. Communication of affect between patient and physician, in *Journal of Health and Social Behavior*, vol. 22, 1981, pp. 18-30.
- [14] Haskell WL. The efficacy and safety of exercise programs in cardiac rehabilitation, in *Medicine & science in sports and exercise*, vol. 26, 1994, pp. 815-823.
- [15] John OP, Donahue EM, Kentle R. The Big-Five Inventory, Technical Report, Insititute of Personality and Social Research, University of California, Berkeley, CA, 1991.
- [16] LeBaron S, Reyher J, Stack JM. Paternalistic vs egalitarian physician styles: The treatment of patients in crisis, in *Journal of Family Practice*, vol. 21, 1985, pp. 56-62.
- [17] Mateas M. An Oz-centric review of interactive drama and believable agents, in MJ Wooldridge & M Veloso (Eds.), *Artificial intelligence today: Recent trends and developments*. NY, Springer, 1999, pp. 297-398.
- [18] Nass C, Moon Y. Machines and mindlessness: Social responses to computers, in *Journal of Social Issues*, vol 56, 2000, pp. 81-103
- [19] Parise S, Kiesler S, Sproull L, Waters K. Cooperating with life-like interface agents, in *Computers in Human Behavior*, vol. 15, 1999, 123-142.
- [20] Raingruber BJ. Settling into and moving in a climate of care: Styles and patterns of interaction between nurse psychotherapists and clients, in *Perspectives in Psychiatric Care*, vol. 37, 2001, pp. 15-27.
- [21] Reeves B, Nass C. *The media equation: How people treat computers, television, and new media like real people and places*, Cambridge University Press, New York, 1996.
- [22] Rousseau D, Hayes-Roth B. Interacting with personality-rich characters, Sept. Technical Report No. KSL 97-06, Knowledge Systems Laboratory, Department of Computer Science, Stanford University, Stanford, CA, 1997.
- [23] Scholl BJ, Tremoulet P. Perceptual causality and animacy. *Trends in Cognitive Sciences*, vol. 4(8), 2000, pp. 299-309
- [24] Siino RM. Making sense of new technology: Organizational sensemaking of autonomous mobile robots as a lead-in to structuring. Unpublished manuscript, WTO, Stanford University, May 2003.
- [25] Thrun S, Beetz M, Bennewitz M, Burgard W, Cremers AB, Dellaert F, Fox D, Haehnel D, Rosenberg C, Roy N, Schulte J, Schulz D. Probabilistic algorithms and the interactive museum tour-guide robot, Minerva, in *International Journal of Robotics Research*, vol. 19, 2000, pp. 972-999.
- [26] Wetter AC, Goldberg JP, King AC, Sigman-Grant M, Baer R, Crayton E, Devine C, Drewnowski A, Dunn A, Johnson G, Pronk N, Saelens B, Snyder D, Novelli P, Walsh K, Warland R. How and why do individuals make food and physical activity choices? *Nutrition Reviews*, vol. 59(3 Pt 2), 2001, pp. 11-20.
- [27] Zebrowitz LA, Collins MA, Dutta R. The relationship between appearance and personality across the life span, in *Personality & Social Psychology Bulletin*, vol. 24, 1998, pp. 736-749.