8-18-2020 2:30 PM

What to Say and How to Say It: the Interplay of Self-Disclosure Depth, Similarity, and Interpersonal Liking in Initial Social Interactions

Yixian Li, *The University of Western Ontario*

Supervisor: Heerey, Erin, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Psychology

Recommended Citation

Li, Yixian, "What to Say and How to Say It: the Interplay of Self-Disclosure Depth, Similarity, and Interpersonal Liking in Initial Social Interactions" (2020). *Electronic Thesis and Dissertation Repository*. 7233.
https://ir.lib.uwo.ca/etd/7233

Abstract

We often initiate social relationships with others through revelations of personal information, or self-disclosure. Self-disclosure is heavily involved in shaping interpersonal liking, but there are disparate and sometimes contradictory findings in the literature regarding the causal relationship between them. Moreover, a lack of careful control in experimental designs in many existing studies failed to eliminate important confounding factors that might provide alternative explanations for the disclosure-liking relationship. Here, we examined the relationships between self-disclosure and interpersonal liking during initial social interactions, while carefully controlling for a potential confounding factor, similarity between the social partners.

Across the first five experiments, I independently manipulated disclosers' self-disclosure depth, i.e., how personal and intimate the disclosures are, and their self-disclosed similarity with their social partners. High self-disclosed similarity was consistently found to lead to greater initial liking of a discloser. In comparison, the experiments failed to find support for the idea that people favor those who self-disclose more deeply, as suggested in the literature. In Experiment 6, I manipulated initial liking within a set of social partners and successfully replicated another disclosure-liking relationship identified in the literature, namely, the effect that people self-disclose to a greater extent to those whom they like. It was also found that, contrary to the expectation, participants' risk-taking tendencies *negatively* predicted their self-disclosure depth to others. In Experiment 7, I extended the investigation to an emerging and novel social context and examined how self-disclosed similarity from an Artificially Intelligent (AI) agent influenced people's perceptions of and responses to the agent. A significant interaction between the perceived identity of the partner (i.e., AI versus human) and level of self-disclosed similarity was found. The results were interpreted in light of the "uncanny valley effect", which suggests that a high level of human realism displayed by an automatic agent could elicit unpleasant or "eerie" feelings.

Through this series of experiments, I iteratively developed the paradigm to more closely mimic real-world social disclosures. The findings help disentangle the causal relationship between self-disclosure and initial liking and provide insights into some of the subtleties and processes underlying relationship formation.

Keywords

Interpersonal liking, self-disclosure, friendship development, first impression, similarity, human-AI interaction

Summary for Lay Audience

Making friends is important. Being able to enjoy good social relationships with other people is beneficial to both our psychological and physical health. The friend-making process frequently starts when we tell each other information about ourselves, such as our past, hobbies, thoughts, and feelings. This act of revealing our own information to another person is called self-disclosure.

What should we self-disclose in our first interaction with another person to best kindle the budding friendship? The past literature suggests that self-disclosing deeper, rather than more superficial, information about yourself might make the other person like you more. However, there are some methodological problems with the previous studies that render this conclusion questionable. Specifically, the fact that you self-disclose more deeply to the other person and that this person likes you more might both have resulted from a greater similarity between the two of you.

To address this issue, we investigated whether self-disclosing more deeply to a stranger makes them like oneself more, after experimentally controlling for the level of similarity between the two people. We consistently found people to like those who self-disclosed a greater similarity to themselves and not those who self-disclosed more deeply to them. In other words, revealing to a stranger that you are similar to them would make them like you more, whereas telling them deeply personal information about yourself would likely not. I also investigated whether the causal relationship is the other way around, namely, whether liking the other person more to start with leads one to self-disclose more deeply to them. My findings supported this account. Finally, we explored whether people like an artificially intelligent (AI) agent more if the agent self-disclosed greater similarity with themselves. Findings suggested that people reacted differently to self-disclosed similarity coming from an AI partner versus a human partner.

Co-Authorship Statement

Dr. Erin Heerey is a co-author for *Does Self-Disclosure Depth Really Matter in Developing Initial Feelings of Liking?* Manuscript submitted for publication. (reported in Chapter 2 of this dissertation). She contributed to the design of the experiments and the editing of the manuscript.

Dr. Erin Heerey and Dr. Jonathan Gratch are co-authors for *Negative Perceptions of a Self-Disclosing AI: the Potential Role of the Uncanny Valley Effect.* Manuscript in preparation. (reported in Chapter 4). Dr. Heerey and Dr. Gratch both contributed to the design of the experiment and the editing of the manuscript. Dr. Gratch provided the computer software, equipment, and lab space for data collection. Dr. Gratch also provided funding for subject recruitment.

Acknowledgments

I would like to express my deep gratitude to my PhD supervisor, Dr. Erin Heerey. Thank you very much for being so responsive, encouraging, supportive, and, when it comes to experimental design, incredibly meticulous. Thank you for always pushing me to learn more and do better. I would not have become the researcher I am today if it were not for you.

I would also like to thank Dr. Richard Sorrentino and Dr. Ross Norman for their guidance in my undergraduate and master programs. Thank you for showing me the entrance to the world of psychological research. It is a wonderful world and I have had quite some fun in it.

Thank you to my supervisory committee members, Dr. Sam Joel and Dr. Bill Fisher, for all your helpful suggestions; thank you to my examination committee members, Dr. Paul Tremblay, Dr. Mike Katchabaw, and Dr. Elizabeth Page-Gould, for your insightful comments and interesting conversations during my defense.

To all my good friends from and outside of graduate school, I thank you whole-heartedly for your love and support. My graduate years has been made so much more enjoyable with all the wine nights, board game nights, dinner hangouts, and fun/nerdy office conversations (especially on renewable energy).

I would like to thank my family in China. Dad, mom, and my granny, I love you very much and probably should express that more often. Thank you for teaching me to be curious, open-minded, persistent, and resilient.

Lastly, thank you very much, Yixiao. This would have been such a lonely and stressful journey if I did not have you and our little family of our dog, bunny, and guinea pig by my side. Thank you for being my charging station whenever my energy is low. You are my partner in this adventure called life and I am in luck.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

1 Theoretical Background and Rationale

Forming strong and meaningful interpersonal relationships is a fundamental human need (Baumeister & Leary, 1995). Positive social relationships constitute one of the dimensions that define people's psychological well-being (e.g., Ryff & Keyes, 1995). Such relationships also promote physical well-being and have even been shown to decrease rates of morality by providing people with social support that buffers against stress and adversity (e.g., Cohen, 2004). In every initial interaction with a stranger lies the opportunity to cultivate a positive and valuable social relationship. Because people form first impressions very quickly (Bar et al., 2006; Willis & Todorov, 2006), which in turn influences both subsequent interactions with others and even long-term social outcomes (e.g., Human, Sandstrom, Biesanz, & Dunn, 2013; Marek, Wanzer, & Knapp, 2004; Sunnafrank & Ramirez, 2004), it is important to identify factors that contribute to interpersonal liking during initial interactions and that help kindle relationship development.

Many of our most rewarding relationships begin when people open up to one another and share their stories, thoughts, and feelings. Self-disclosure, the act of revealing personal information to another (Collins & Miller, 1994), has therefore been identified as crucial to the development and maintenance of interpersonal relationships (Altman & Taylor, 1973; Bauminger et al., 2008; Berger & Calabrese, 1975; Keelan et al., 1998; Kerr et al., 1999; Reis & Shaver, 1988). Understanding how self-disclosures contribute to initial liking can guide people's decisions on what and how to self-disclose when they first meet another person to best nourish budding interpersonal relationships. The series of experiments presented in this dissertation thus examined how self-disclosures shape initial liking of the discloser (Experiments 1-5) and people's perceptions of a self-disclosing artificially intelligent (AI) agent (Experiment 7). We also examined how people's liking of another person and their risk-taking tendency influence their own level of self-disclosure to that person (Experiment 6).

1.1 Related theories

1.1.1 Social penetration theory

One primary theoretical framework for understanding the role of self-disclosures in interpersonal interactions is social penetration theory (Altman & Taylor, 1973). This theory distinguishes between the *breadth* and the *depth* of self-disclosures. The breadth of self-disclosures refers to the range of topics that people disclose, whereas disclosure depth characterizes the degree to which a disclosure is personal or intimate. Using the metaphor of an "onion", Altman & Taylor (1973) liken disclosures to a series of nested layers – with the outer, peripheral layers representing more superficial "surface" level disclosures and inner, central layers representing more personal, intimate or private information. In addition, there are different disclosure content categories or major topic areas, forming "wedge-like" shapes; within each category or wedge there are different layers of disclosure depth. See Figure 1-1 for a visual illustration of this structure.

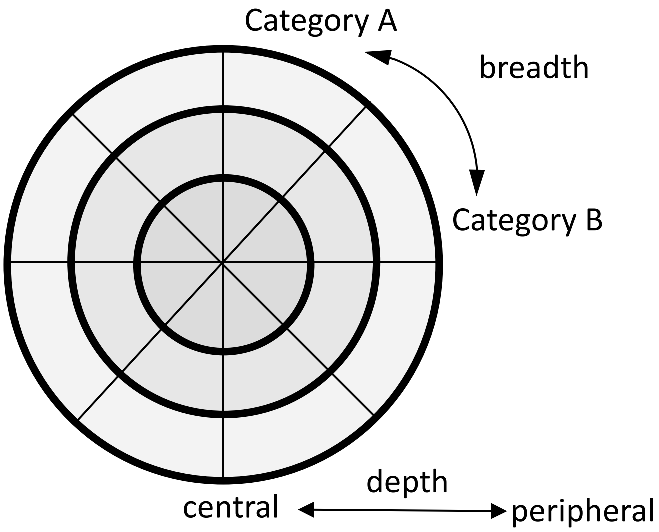


Figure 1-1 Structure for disclosure depth and disclosure breadth proposed in social penetration theory. Adapted from Altman and Taylor (1973).

According to social penetration theory, a few patterns characterize the process in which people change their breath and depth of self-disclosures as a relationship deepens (Altman & Taylor, 1973; Derlega, 2009). First, people initiate self-disclosure by revealing superficial information about themselves. As a relationship develops, they gradually begin to disclose more personal and intimate information. The decision to increase the intimacy and depth of one's self-disclosure to another is influenced by people's assessment of the reward (e.g., agreement and approval of the receiver) versus the cost (e.g., possible disagreement and social rejection) ratio of the previous self-disclosures (Taylor & Altman, 1975). Second, self-disclosures can be organized by "breadth category", which refers to different topic areas (e.g., family, hobbies, religion, etc.). As a relationship develops, people not only move to self-disclose more deeply within the same breadth category, but also expand disclosure breadth by disclosing information in categories related to those in which disclosures have been exchanged and certain level of intimacy has been achieved. Third, once a category or an "inner layer" is "unlocked" in a conversation and the corresponding self-disclosures are met with favorable outcomes, these topics can be revisited in future exchanges. People can also refrain from self-disclosing information in certain topics for reasons such as the topics being too private or taboo, and therefore "deny access" about these areas of self to another person.

1.1.2 Intimacy as an interpersonal process

Another theoretical framework on how self-disclosures are involved in relationship development is the transitional model of intimacy proposed by Reis and Shaver (1988). In comparison to social penetration theory, Reis and Shaver's model further focuses on the dynamic process between the discloser and the receiver. According to this model (Figure 1-1), the interaction starts with one's self-disclosure to a social partner, which is affected by a variety of factors such as the discloser's motives, needs, and fears. For example, a desire for affection and understanding would promote self-disclosure whereas a fear of exposure and abandonment would make someone reluctant to share certain information. Next, a receiver interprets and responds to the disclosures; both their interpretations and responses are influenced by their own motives and needs. Positive responses to the self-disclosure would help promote a sense of connectedness whereas negative responses or deliberate nonresponses might keep the interaction from becoming more intimate. Finally, the initial discloser interprets the receiver's responses. To experience the interaction as intimate, the discloser must perceive the receiver's responses to their initial self-disclosure as understanding, validating, and caring, which might, in return, promote further exchange and expressions of feelings between the two people. Interestingly, the authors asserted that revealing one's emotions and feelings is more important to the development of intimacy than sharing "merely facts", as the former allows the receiver to respond and validate the "inner self" of the discloser, which constitutes an affective core that persists across the lifespan. The concept of "inner self" is comparable to the "inner layers" of personality proposed in social penetration theory.

Reis and Shaver's model, therefore, stresses the importance of deep self-disclosures in relationship development, as does social penetration theory.

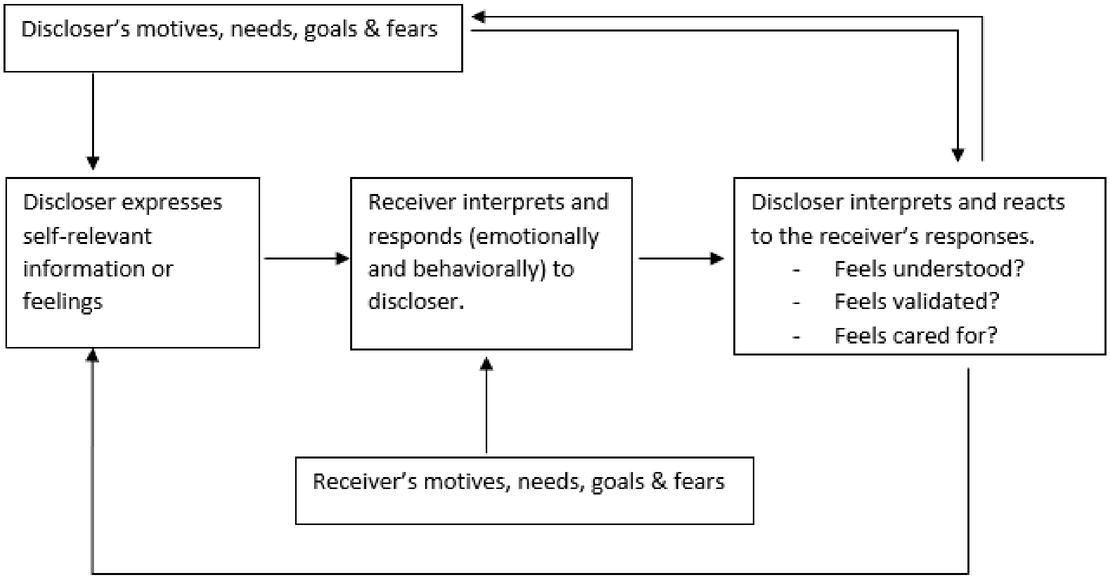

Discloser's motives, needs, goals & fears
Discloser expresses self-relevant information or feelings
Receiver interprets and responds (emotionally and behaviorally) to discloser.
Discloser interprets and reacts to the receiver's responses.
- Feels understood?
- Feels validated?
- Feels cared for?
Receiver's motives, needs, goals & fears

Figure 1-2.Transactional model of intimacy process. Figure adapted from Reis and Shaver (1988).

1.2 Empirical evidence

Collins and Miller (1994) conducted a thorough and systematic literature review on the empirical studies that examined the link between self-disclosure and interpersonal liking published between 1955 and 1992. Particularly, they pointed out that the there are three distinct, although often related, disclosure-liking relationships that were not clearly distinguished in the literature, which might have contributed the mixed findings in the literature. The three effects that Collins and Miller investigated included (1) whether people would like those who self-disclose at a more intimate rather than superficial level to them , (2) whether people would disclose more to those whom they like, and (3) whether the act of self-disclosing to someone would enhance the discloser's liking of the receiver). The authors concluded that the evidence supported significant and positive disclosure-liking relationships regarding all the three effects.

The first effect reviewed in Collins and Miller, "people develop more interpersonal liking for others who self-disclose more intimately to them", seemed to receive most research interest among the three. The authors included 55 studies, reporting a total of 94 effect sizes, that provided evidence on whether self-disclosures lead to greater interpersonal liking. Following the logic of social penetration theory, the authors further separated studies that manipulated or measured disclosure *depth* versus disclosure *breadth*. It was found that studies that operationalized the level of self-disclosure in terms of the intimacy or depth of the disclosures induced significantly stronger effects on liking than those that operationalized it as the sheer quantity or breadth of the disclosures. As a result, they concluded that the empirical evidence is in support for the effect that people like those who self-disclose more intimately rather than superficially to themselves. The authors provided a few potential explanations for why more intimate self-disclosures might lead to greater liking of the discloser. First, deep self-disclosures may signal the discloser's liking and desire for a more intimate relationship with the receiver, which can be viewed as rewarding by the receiver and thus leads to greater liking of the discloser. Second, a receiver might form positive beliefs about a discloser (e.g., trusting and warm), if the discloser reveals more intimate and potentially vulnerable information about themselves. These positive beliefs would then in turn lead to greater liking of the discloser. Interestingly, despite the general conclusion that deeper self-disclosures lead to greater liking of the discloser, there was wide variation in not only the magnitude of the effect but also direction of the effect across studies. The authors suggested that this variation was partially related to the different study methodologies. Specifically, the correlational studies generated the strongest positive effects (i.e., Cohen's d=0.845), lab-based experiments tended to generate weaker effects (i.e., Cohen's d ranging from 0.191 to 0.378), and field studies where a confederate interacted and self-disclosed to a stranger generated a *negative* effect (i.e., Cohen's d= -0.308).

The second effect reviewed in Collins and Miller, “people tend to self-disclose more to those whom they like more”, received slightly less research interest than the first. The authors identified 22 studies that reported a total of 31 effect sizes regarding this effect. However, the mean effect size for this effect (Cohen’s d=.717) was considerably larger than that for the first effect (that people tend to like those who disclose more to them; Cohen’s d=.281). As with the first effect, the effect size for the second effect also varied across studies that used different methods, with the correlational studies generating stronger effects (Cohen’s d=1.105) compared to the lab-based experiments (Cohen’s d= .277 to .449). But unlike the first effect, there was mostly a consensus among the reviewed studies on the positive direction of the effect: 90% of effect sizes (i.e., 28 out of the 31) included in the review regarding this effect were positive, compared to the 67% of effect sizes (63 out of the 94) reviewed regarding the first effect that were positive. Thus, the effect that people self-disclose to a greater extent to those whom they like more seems to be overly supported by empirical evidence. Social penetration theory (Altman & Taylor, 1973) and the transactional model of intimacy (Reis & Shaver, 1988) both suggest that people’s self-disclosures are influenced by the anticipated outcomes of the disclosure. The anticipated reward, such as approval, understanding, and validation from a liked other might thus motivate people to engage in greater self-disclosure to a liked social partner.

Only five empirical studies (and five effect sizes) were included in Collins and Miller that provided evidence regarding the third effect, namely, the act of self-disclosing increases the discloser’s liking of the receiver. Again, there was a significant variation across the studies such that two generated large and positive effect sizes whereas the remaining three had effect sizes of zero. In addition, as the authors pointed out, research examining potential mechanisms underlying this effect also generated mixed results: while some researchers suggested that self-disclosures are personally rewarding and cathartic for the discloser (Jourard, 1959), especially those with traumatic experiences (Pennebaker, 1985), others suggested that self-disclosures might make people aware of the negative discrepancies between ideal and actual self, resulting in negative affect (Archer et al., 1982). The cumulative evidence regarding this effect thus seems to be weak.

Since the beginning of this century, self-disclosure research has largely focused on examining the role of self-disclosures in computer-mediated communications and in online social relationship formation. These studies have generally found similar effects as in face-to-face social interactions. For example, greater online self-disclosure appears to increase liking from interaction partners (Kashian, Jang, Shin, Dai, & Walther, 2017; Utz, 2015), enhances perceived online social support (Lee, Noh, & Koo, 2013), and generates more positive relationship outcomes (Gibbs, Ellison, & Heino, 2006; Yum & Hara, 2005). Similarly, as in face-to-face interactions, receivers' responses to one's self-disclosures contribute to interaction outcomes in computer-mediated communications (Dai, Shin, Kashian, Jang, & Walther, 2015).

1.3 Do deep self-disclosures lead to greater initial liking?

Previous work highlights the importance of deep or intimate self-disclosures in shaping interpersonal relationships (Collins & Miller, 1994). As previously discussed, Reis and Shaver (1988) argued that this may be because deeper self-disclosures, such as those about emotions, are more important to relationship development than the more superficial disclosures of self-relevant facts, as the former allow a receiver to understand the discloser's "emotional core" and thus allow the interaction to become more intimate. This echoes the process described in social penetration theory (Altman & Taylor, 1973) in which two people gradually reveal information pertaining to the "inner layers" of personality such as core values and fears; here, the increased disclosure-depth may be viewed as a barometer of closeness. However, the processes described in these theoretical frameworks might be more readily applied to long-term and ongoing relationships where two people have a basic understanding of one another and expect repeated interaction.

Interestingly, Collins and Miller found support for the positive effects of deep self-disclosures on interpersonal liking for studies that used a "get-acquainted" paradigm, where participants interacted with another participant or a confederate for the first time. This suggests that deep self-disclosures might indeed lead to positive social outcomes not only in ongoing relationships, but also in initial social interactions. There were, however, a few substantial methodological limitations in these studies reviewed in Collins and Miller that hinder our confidence in the positive effect of self-disclosure depth on liking during initial social interactions.

First, many of these studies had very small sample sizes. Among the studies reviewed in Collins and Miller, many had 10 or fewer participants per condition; some had fewer than five per condition. Small sample sizes contribute to low statistical power, which in turn reduces the likelihood that a statistically significant result reflects an actual true effect (i.e., induces a higher false positive rate) (e.g., Button et al., 2013).

Second, many of the experimental studies reviewed used confederates who changed their behaviours when interacting with participants depending on the specific manipulation conditions. As some of these studies acknowledged (e.g., Archer & Berg, 1978), the confederates were usually aware of the predictions, which might have inadvertently biased their behaviours, and in turn biased participants' responses to them (see Holman, Head, Lanfear, & Jennions, 2015; Wicherts et al., 2016).

Finally, and most importantly, the interactions that take place between participants and confederates or between pairs of participants are difficult to carefully control (Kuhlen & Brennan, 2013). As a result, even though the researchers manipulated and/or measured disclosure depth, these measurements are most likely not independent from other confounding factors that might provide alternative explanations to the observed results. For example, when two participants freely interact, one might receive deep self-disclosures from the other and self-report greater liking of that person, but not because the former led to the latter. Rather, both these events could have resulted from a positive interaction experience. Or, when interacting with a confederate who self-disclosed intimately, a participant might have liked the confederate more not because of the confederate's deep disclosures per se but rather because the participant reciprocated the confederate’s disclosures and the confederate responded more positively to the participant in return. Thus, the subtle tone and affect differences participants experience when confederates are aware of task conditions might inadvertently cultivate a friendlier environment. Without the independent manipulation of self-disclosure depth while carefully controlling potential confounding factors, a causal relationship between self-disclosures and liking for the discloser cannot be established.

1.4 Similarity as a confound in the disclosure-liking link

One alternate factor that could greatly influence the experience of initial interactions, and thus confound the effects of self-disclosure on initial liking, is the level of similarity between interaction partners. This similarity-attraction effect (Byrne, 1971) exerts that people tend to like those whom they perceive as more similar to themselves. This effect has received widespread empirical support (e.g., Montoya & Horton, 2013; Montoya, Horton, & Kirchner, 2008; Tidwell, Eastwick, & Finkel, 2013). For example, in an extensive meta-analysis that included over 300 empirical studies involving more than 35,000 participants, Montoya et al. (2008) found moderate-to-large effects of both actual similarity (r=.47) and perceived similarity (r=.39) on interpersonal attraction.

Why does similarity lead to attraction? One explanation for the similarity-attraction effect was built on the *information processing* approach to interpersonal attraction (Ajzen, 1977), which asserts that people form their impressions of and attraction to another based on the information they have about the other person. Positive and favorable information about the other person leads to attraction to that person. As we tend to view our own characteristics positively, we also view others who share these characteristics positively, which leads to greater attraction to them (Ajzen, 1974; Montoya & Horton, 2013). Moreover, people perceive similar others as socially warmer and more intellectual (Lydon et al., 1988), which would also lead to greater attraction to them. Other researchers have suggested that similarity attracts because we expect the similar other to like us, which in turn leads to our greater attraction to them (Aronson & Worchel, 1966; Condon & Crano, 1988; Insko & et al, 1973).

Byrne and colleagues suggested a reinforcement-affect model of interpersonal attraction (Byrne, 1971, 1997; Clore & Byrne, 1974), which is an overarching theoretical account that can potentially incorporate all the aforementioned explanations for the similarity-attraction effect. This model asserts that one's attraction to another is proportionally reinforced by their positive affect, relative to negative affect, associated with the target person. Similarity is an example of the stimuli that can elicit one's positive affective responses and thus reinforce their attraction to another. Similarity may induce such positive affect because similar others validate people's attitudes and world-views, which satisfy their need to understand, evaluate, and predict their environment (Byrne & Clore, 1967). It may also induce positive affect through other routes as previously described, such as by contributing to a more positive evaluation of the other person (e.g., Ajzen, 1974) or by allowing us to infer positive evaluations *of us* by others (e.g., Condon & Crano, 1988).

Similarity might seriously confound the positive effect of self-disclosure depth on interpersonal liking. As reviewed and supported in Collins and Miller (1994), people do not only like those who self-disclose more (the first effect), but also themselves disclose more to those whom they like (the second effect). It is therefore quite likely that if there is a high level of similarity between the discloser and the receiver, they will like each other more, resulting in greater exchange of self-disclosures with each other. The effects of deep self-disclosures on interpersonal liking might therefore be, at least in part, by-products of perceived similarity between interaction partners. This might be especially true in studies where participants were given some time to engage in free conversation with other participants or the confederate.

Another possibility is that similarity may be a potential *enabling factor* for the effects of disclosure depth on liking. As previously mentioned, deep self-disclosures might lead to greater liking of the discloser because they signal the discloser's interest in a more intimate social relationship with the receiver. It is likely that the receiver would welcome such attention more if they also like the discloser. Therefore, deeper self-disclosures might only lead to greater liking when the receiver likes the discloser to start with, possibly as a result of the greater perceived similarity with the discloser.

As previously mentioned, Collins and Miller (1994) found considerate variations in the magnitude and directions among studies that examined the effect that people like those who self-disclose more intimately to them more. Similarity between social partners might help explain such variations. Specifically, the authors found that the correlational studies generated the strongest positive effect, followed by lab-based experiments, whereas the field studies generated a negative effect. Considering the likely pre-existing similarity between the social partners in the different types of studies, the similarity-attraction effect predicts exactly this pattern of variation. Specifically, it is likely that people in existing and ongoing social relationships are more similar, or at least perceive each other to be more similar, than those randomly paired in experiments. Likewise, most lab-based studies used university student samples while participants in the field studies were members of the general public, with arguably less in common than university students living in the same geographic region. Participants in the lab-based experiments might thus perceive greater similarity with each other or with a student confederate, compared to the participants in the field studies. The greater similarity or perceived similarity between social partners in the correlational studies and lab-based experiments might therefore lead to greater initial liking for disclosers, enabling the positive effects of self-disclosure depth on liking. In comparison, a lack of perceived similarity might limit the degree to which the receiver liked and welcomed deep self-disclosures from the confederate in the field studies, resulting in a negative effect of disclosure depth on liking of the discloser.

1.5 The current experiments

Without carefully controlling similarity between the two partners, one cannot confidently predict the role of self-disclosure depth in shaping initial social outcomes. The current experiments thus examined how self-disclosure depth is linked to interpersonal liking in initial interactions, independent of the effect of similarity between the social partners. Specifically, in Experiments 1 to 5 presented in Chapter 2, we independently manipulated self-disclosure depth and similarity and examined their effects on liking of the discloser in initial social interactions. In Experiment 6, presented in Chapter 3, I tried to replicate Collins and Miller's (1994) assertion that liking of a partner predicts the extent to which participants decide to self-disclose to the partner. I also examined how individual differences in factors such as risk-taking tendencies influence people's self-disclosure decisions. In Experiment 7, which was present in Chapter 4, we extended our investigation to an emerging social context, the interactions between a social AI agent and its human user. Specifically, we examined how self-disclosed similarity from an AI influences the user's perceptions of and preferences for it.

The experiments were designed such that the methodological limitations as previously discussed could be adequately addressed. First, most of the experiments used a within-subject design. This within-subject design allows each participant to experience all our manipulation conditions, which reduced noise and increased our statistical power. As a result, we were more confident in both our ability to detect the proposed effect as well as the validity of any significant effects that we did find. Second, to address the concerns associated with using a confederate, participants completed the entire experiment on a computer and interacted with computerized avatars whom they believed were other participants in all our experiments except for Experiment 7. This effectively eliminated experimenter bias and the possibility that participants' behaviours might have been influenced by certain characteristics of confederate behaviour. In Experiment 7, although the nature of the task did not allow a within-subject design, careful randomization was used to reduce potential confounds. Measures were also taken to record and assess experimenter bias. Finally, a priori power analyses helped to ensure that we recruited adequate participant samples to enhance replicability.

Chapter 2

2 Experiments 1 to 5: Does Disclosure Depth Matter?

In Experiments 1 to 3, we independently manipulated the levels of disclosure depth and similarity between disclosers and receivers to examine the unique effects of disclosure depth on interaction outcomes. In Experiment 4, in addition to similarity and disclosure depth, we manipulated the level of "personalistic" attribution, or the extent to which participants believed that another's self-disclosures were selectively made to them. Finally, in Experiment 5, we examined how similarity and the *reciprocity* of disclosure depth influenced interpersonal liking in initial interactions. Except for the first experiment, we pre-registered the methods and hypotheses of all the experiments[1]; all the datasets and R scripts for data analyses reported in this paper are publicly available[2]. We used the same general experimental paradigm throughout all the experiments, with slight modifications in each.

2.1 General methods

Analogous to animals' foraging behaviours in search of food, people seek out and gather information available in the environment that helps reduce uncertainty and allows them to obtain greater rewards (Abram et al., 2016; Manohar & Husain, 2013). In a similar manner, people might seek out others' self-disclosures to reduce uncertainty within the social environment and achieve social rewards, such as affiliation with liked others. We therefore devised an experimental paradigm in which participants were free to explore and seek out others' self-disclosures while allowing us to systematically manipulate the information to which participants were exposed.

[1] Pre-registrations: Experiment 2: https://osf.io/fqz62; Experiment 3: https://osf.io/d6ce8 ; Experiment 4: https://osf.io/xdreb; Experiment5: https://osf.io/e6xcg.

[2] Data and R scripts storage: Experiment 1- 2: https://osf.io/utuvp/; Experiment 3-5: https://osf.io/pc2bn/.

2.1.1 Material preparation

During the stimulus development phase of this work, we generated 40 multiple-choice questions on topics that varied in the degree to which they represented deep-level disclosures. These 40 multiple-choice questions and answers were subsequently used as self-disclosure items in our experimental paradigm, allowing us to manipulate both similarity and disclosure depth independently. One-hundred thirty-nine university students (37 male, 102 female) aged from 17 to 35 (*M*=18.9, *SD*=2.1) answered 70 open-ended get-to-know-you questions (e.g., “What do you like to do in your free time?”; “What do you want to change the most about yourself?”). For each of these questions, participants also rated how personal they felt the question was on scale of 1 (“not at all personal”) to 7 (“very personal”). We selected 20 questions that were rated as relatively low in how personal they were (*M*=2.24, *SD*=0.35) as low disclosure-depth topics and another 20 that were rated as relatively highly personal (*M*=4.02, *SD*=0.43) as high disclosure-depth topics. We deliberately chose topics that were considered appropriate in their level of intimacy for initial social interactions among university students, as evidence suggests that overly intimate self-disclosures from strangers can elicit negative responses (e.g., Caltabiano & Smithson, 1983; Chaikin & Derlega, 1974). For each of the 40 disclosure topics, we generated four multiple-choice answers based on the most common themes in participants’ open-ended answers (Appendix A).

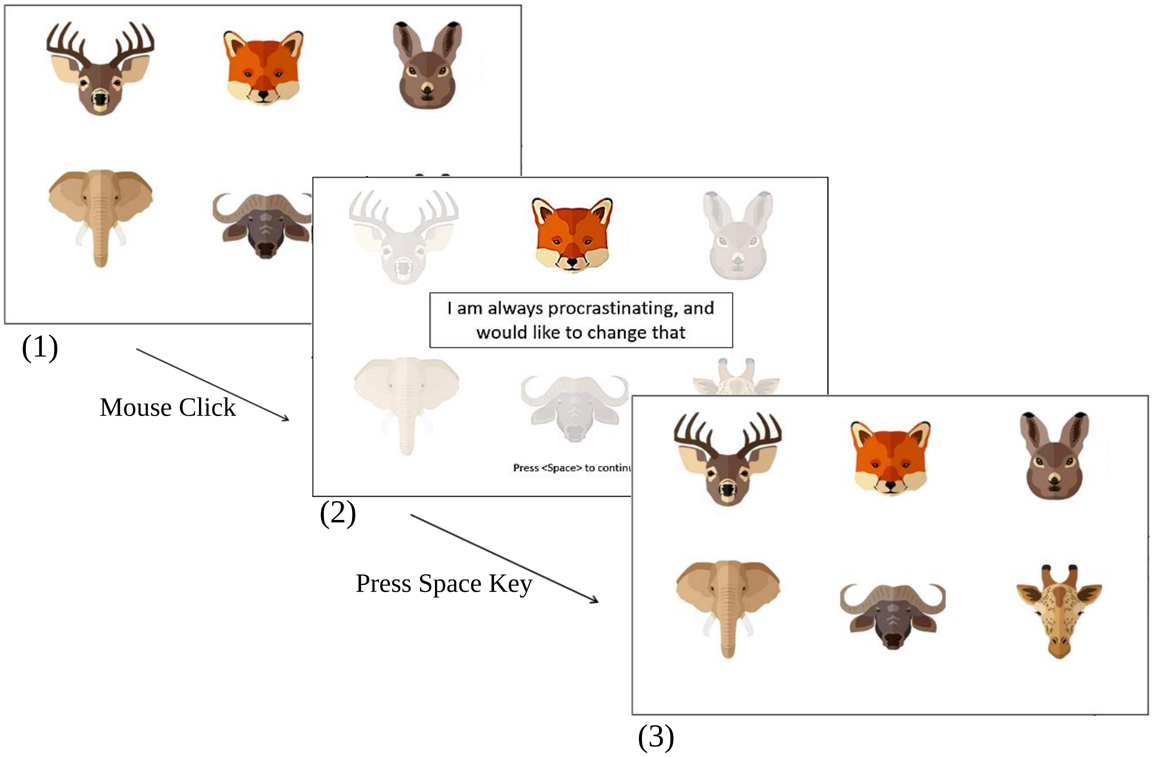
2.1.2 Experimental paradigm

In all the studies, participants arrived at the lab in groups, even though they completed the task in individual lab rooms on a computer. The computer task began by first allowing participants to select an avatar picture that they believed would represent them during the task. Participants then gave their own responses to each of the 40 multiple-choice questions. Next, participants were allowed to “get to know” the other participants in a virtual “social environment” where six avatars, each representing one participant, were displayed on the computer screen. To gain information about the other players, they clicked on another player's avatar. The computer then provided them with a disclosure from that avatar. Each click on an avatar earned participants access to one piece of information about that avatar. Each avatar provided 20 unique disclosures that met the restrictions specific to the relevant manipulations (described below) in a random order. If participants clicked an avatar more than 20 times, the computer re-randomized the disclosure deck and participants viewed repeated information. See Figure 2-1 for an illustration of the process[3]. Participants had 6 minutes to learn about the avatars and were instructed to get to know the other players as well as possible during that time. Participants sampled the social environment freely and click-order was not enforced during that time.

In Experiments 1 to 3, the computer manipulated the six avatars' disclosures depending on (1) similarity to participants' responses to the initial 40 questions and (2) on the pre-rated disclosure depth. Using a factorial design, three levels of similarity (high, medium, and low) were fully crossed with two levels of disclosure depth (high and low). That is, each avatar represented one of the six manipulation conditions, reflected in the 3X2 within-subjects design. We manipulated similarity by changing the frequency with which each avatar disclosed information that matched the participant's own disclosure. Specifically, 80% of the 20 possible disclosures from the high-similarity avatars were the same as the participant's own choices, while the remaining 20% were different. In comparison, 50% and 20%, respectively, of the disclosures from medium- and low-similarity avatars were the same as the participant's own answers. To manipulate disclosure depth, 80% of the disclosures from high-depth avatars were high-depth items and the remaining 20% were low-depth items. In comparison, 80% of the disclosures from low-disclosure-depth avatars were low-depth items and the remaining 20% were high-depth items. In Experiments 4 and 5, the manipulations were slightly different, and

[3] The avatar pictures used in Experiment 1were anonymous cartoon headshots. For all the remaining experiments, animal pictures as shown in Figure 1 were used to avoid conveying any information regarding gender or ethnicity that might bias participants' responses.

are described in later sections. After the disclosure phase, the computer automatically advanced to the dependent measures. Upon completing the dependent measures, participants were debriefed, probed for any suspicion about the manipulation and deception, and thanked for their participation.



I am always procrastinating, and would like to change that

Figure 2-1. Participants were presented with the avatar pictures (1). They were able to click on any avatar to view a self-disclosure from that avatar (2). After 2 seconds, pressing the Space key then returned them to the selection screen (3).

2.1.3 Core dependent measures

The core dependent measures in the current experiments were a list of questions assessing participants' perceived similarity with the avatars ("The degree to which they are similar to you"), their knowledge of the avatars ("How well you think you've gotten to know them"), preferences for the avatars ("How comfortable you'd feel asking them for advice", "How much you'd like to admit them to your circle of friends", "How much you'd like to actually meet them"), and trait perceptions of the avatars ("How trustworthy you think they are", "How friendly you think they are"). The preference questions were partly adapted from items use in Coyne (1976) that measure people's willingness to engage in future interactions with a target person. In Experiments 1 to 3, these items were presented to the participants in the format of ranking questions, in which participants ranked all the avatars in relation to each other on each item. We used ranking questions to measure participants' experiences of the avatars because the force-choice format of the ranking items allowed us to examine relative preferences for the avatars in terms of avatar characteristics. In Experiments 4 and 5, participants rated each avatar on each item using 7-point Likert scales, allowing us to aggregate the items to examine how our manipulations influenced participants' overall preferences for the avatars instead of looking at each item individually. Additional dependent measures in several studies are described in their corresponding sections.

In each of these experiments, participants consented to the experiment procedures before beginning the experiment and, because of the deception involved, documented their fully informed consent during debriefing. The University's Nonmedical Research Ethics Board approved all experimental procedures.

2.1.4 General hypotheses

The main hypotheses were the same in most of these experiments.

H1. We expected a main effect of similarity, such that higher levels of similarity would lead to greater perceived similarity, higher social preferences, and more positive trait perceptions of the avatars.

H2. We expected a main effect of self-disclosure depth, such that higher levels of disclosure depth would lead to greater perceived knowledge, higher social preferences, and more positive trait perceptions of the avatars.

H3. Following the argument that deeper self-disclosures provide information regarding one’s “inner layers of personality” that are important to their self-concept (Altman & Taylor, 1973), the recipient of the disclosures might thus like the discloser even more when their similarity was perceived to be at a deeper rather than more superficial level. As a result, we expected similarity and self-disclosure depth to interact with each other in influencing initial liking of the discloser, such that participants would show greater social preference and more positive trait perceptions of the high-disclosure-depth avatars than the low-disclosure-depth avatars to a greater degree when those avatars displayed higher rather than lower similarity to themselves.

Additional hypotheses in Experiments 4 and 5 are described in their corresponding sections.

2.1.5 Overview of the experiments

To make the procedure more closely mimic real social interactions, we made slight modifications to the experimental paradigm in each iteration of the experiment. See Table for a brief description of each experiment. The detailed changes are described in the corresponding sections.

Table 1 Brief Descriptions of Experiments 1 to 5

Experiment	Brief Description
1	Participants were instructed that the self-disclosures that they read came from previous participants.
2	Participants were instructed that the self-disclosures that they read were from the other participants from the same session, and they expected to engage in face-to-face interactions with some of those others later in the experiment.
3	In addition to the set up in Experiment 2, participants were told to choose which of their own answers to share with each avatar. This propagated the belief that the self-disclosures they read were intentionally shared with them by others.
4	In addition to the set up in Experiment 3, participants were randomly assigned to one of two between-subject conditions that varied in the proportion of self-disclosures that were said to have been made to them only and not to other participants (i.e., personalistic attribution: high vs. low).
5	Participants engaged in “back-and-forth” exchanges of self-disclosures with each avatar. These exchanges varied in disclosure similarity (high vs. low) and how frequently avatars reciprocated a participant’s self-disclosure depth (i.e., reciprocity of depth: high vs. medium vs. low).

2.2 Experiment 1

The goal of Experiment 1 was to examine baseline effects of similarity and disclosure depth in a relatively asocial context to learn whether there are intrinsic effects associated with discovering interpersonal similarity or receiving relatively high-depth personal information. For this reason, participants learned that the “people” they would get to know about in the disclosure phase were previous participants in the experiment, whose self-disclosures were randomly chosen by the computer when participants clicked on their profile pictures. Because this experiment was highly exploratory, we did not preregister any predictions. Instead, we wanted to simply determine whether similarity and disclosure depth would shape participants' responses to the avatars.

2.2.1 Participants

A total of 168 university students took part in the experiment in exchange for partial course credit. We excluded seven participants for inattentive responding, as their total number of clicks on the avatars was below two standard deviation of the mean number of clicks across all participants (*M*=69.82, *SD*=15.17). The remaining 161 participants (75 male, 84 female, 2 not specified) aged from 17 to 23 years (*M*=18.69, *SD*=0.95). Seventy-eight (48%) were Caucasian, 47 (29%) were Asian, 12 (7%) were Middle Eastern, 5 (3%) were people of African descent, 14 (9%) were of other or mixed ethnicities, and the rest 5 (3%) did not respond to the item requesting this information.

2.2.2 Data analysis

Due to their ordinal nature, the ranking responses for our dependent measures were analyzed using cumulative link models (i.e., ordinal regression models) with the *ordinal* package (Christensen, 2015) in R (version 3.6.3). A cumulative link model is a model for ordinal-scale observations. It links, or transforms, the cumulative probabilities of a response falling in an ordered category or below to a linear function of the predictors, thus allowing us to estimate effects of the predictors on the transformed cumulative probabilities (Christensen, 2019).

To test our hypotheses, we used the *clm* function in the *ordinal* package to fit the cumulative link model to the ranking responses for each question. We included similarity, disclosure depth, and their interaction as the predictors in the models. A global sum-to-zero contrast was set to compare different manipulation levels to the grand mean. Because we expected a linear effect of our similarity manipulations, the medium-similarity level was coded as -1 to provide coefficients for the high- and low-similarity levels. Here, we first report the results of Type III Analysis of Deviance (ANODE) based on Wald χ^2- tests for any significant overall effects of similarity, disclosure depth, and their interaction (Christensen, 2019). ANODE tests the significance of an effect by examining the change in deviance (i.e., the goodness-of-fit chi-square value) when the effect is added to the model. We then report the specific coefficients associated with the predictors, which reflect the direction and magnitude of the effects. For brevity, we only report statistics for the significant effects in the text. The full details of the models and coefficients are reported in Appendix B.

In addition to similarity and disclosure depth, we also included the number of times that a participant clicked on each avatar as an additional predictor in the models. This allowed us to control for any confounding effects of participants' knowledge of a specific avatar on their liking of that avatar. The number of clicks was significantly and positively associated with greater preferences for an avatar across all the ranking questions. For brevity, we do not report these effects in the text as they are not of primary interest. Again, we report full details of all models and their estimated effects in Appendix B.

2.2.3 Results

Consistent effects were observed across all our ranking items. We found a significant effect of similarity but no effect of disclosure depth or any interactions on all dependent measures. The Analysis of Deviance using Type III Wald chi-square tests suggested that similarity significantly predicted participants' perceived similarity, $\chi^2(2)=26.82$, p <.001, and perceived knowledge of the avatars, $\chi^2(2)=17.61$, p <.001. Specifically, participants were more likely to rank the high-similarity avatars as being more similar to them, b=0.536, SE=0.082, p<.001, and the ones that they had gotten to know better, b=0.423, SE=0.083, p<.001. These coefficients correspond to odds ratios of 1.71 and 1.53, respectively. That is, the odds of the high-similarity avatars being ranked higher on perceived similarity were 1.71 times, and on perceived knowledge 1.53 times, that of an "average" avatar. Conversely, the low-similarity avatars were less likely to be ranked higher on either perceived similarity, b= -0.471, SE=0.081, p<.001, or perceived knowledge, *b*= -0.269, *SE*=0.081, *p*<.001. These coefficients correspond to odds ratios of 0.62 and 0.76, respectively. That is, the odds of the low-similarity avatars being ranked higher on perceived similarity were 0.62 times, and on perceived knowledge 0.76 times, that of an "average" avatar. No significant effects of disclosure depth or any interaction were found on either perceived similarity or perceived knowledge. See Figure 2-2 for the proportions of participants that placed each of the avatar at each rank place for each of the ranking measures. For brevity, we report the corresponding odds ratios for the coefficients from here onwards in Appendix B instead of in the text.

Similar effects were observed for the preferences items. There was a significant effect of similarity on how much the participants felt comfortable asking the avatars for advice, $\chi^2(2)$=21.74, *p*<.001, how much they'd like to admit the avatars to their circle of friends, $\chi^2(2)$=29.57, *p*<.001, and how much they'd like to actually meet the avatars, $\chi^2(2)$=25.29, *p*<.001. The high-similarity avatars were more likely to be ranked higher on all the three items (*b*=0.360, *SE*=0.081, *p*<.001; *b*=0.551, *SE*=0.081, *p*<.001; and *b*= 0.610, *SE*=0.083, *p*<.001, respectively in the order of the questions) and the low avatars were less likely to be ranked higher on any of the three items (*b*=-0.432, *SE*=0.081, *p*<.001; *b*=-0.444, *SE*=0.081, *p*<.001; *b*=-0.499, *SE*=0.081, *p*<.001, respectively in the order of the questions). Again, no effects of disclosure depth or an interaction were found on any of the social preferences items.

Finally, a significant effect of similarity was observed in participants' rankings of perceived trustworthiness, $\chi^2(2)$=10.54, *p*=.005, as well as avatar friendliness, $\chi^2(2)$=14.85, *p*<.001. As above, the high-similarity avatars were more likely to be ranked higher on both perceived trustworthiness, *b*=0.278, *SE*=0.081, *p*<.001, and friendliness, *b*=0.369, *SE*=0.082, *p*<.001. The low-similarity avatars, on the contrary, were less likely to be ranked higher on either trustworthiness, *b*=-0.176, *SE*=0.080, *p*=.028, or friendliness, *b*=-0.290, *SE*=0.080, *p*<.001. We found no effects of disclosure depth or an interaction on any of these items.

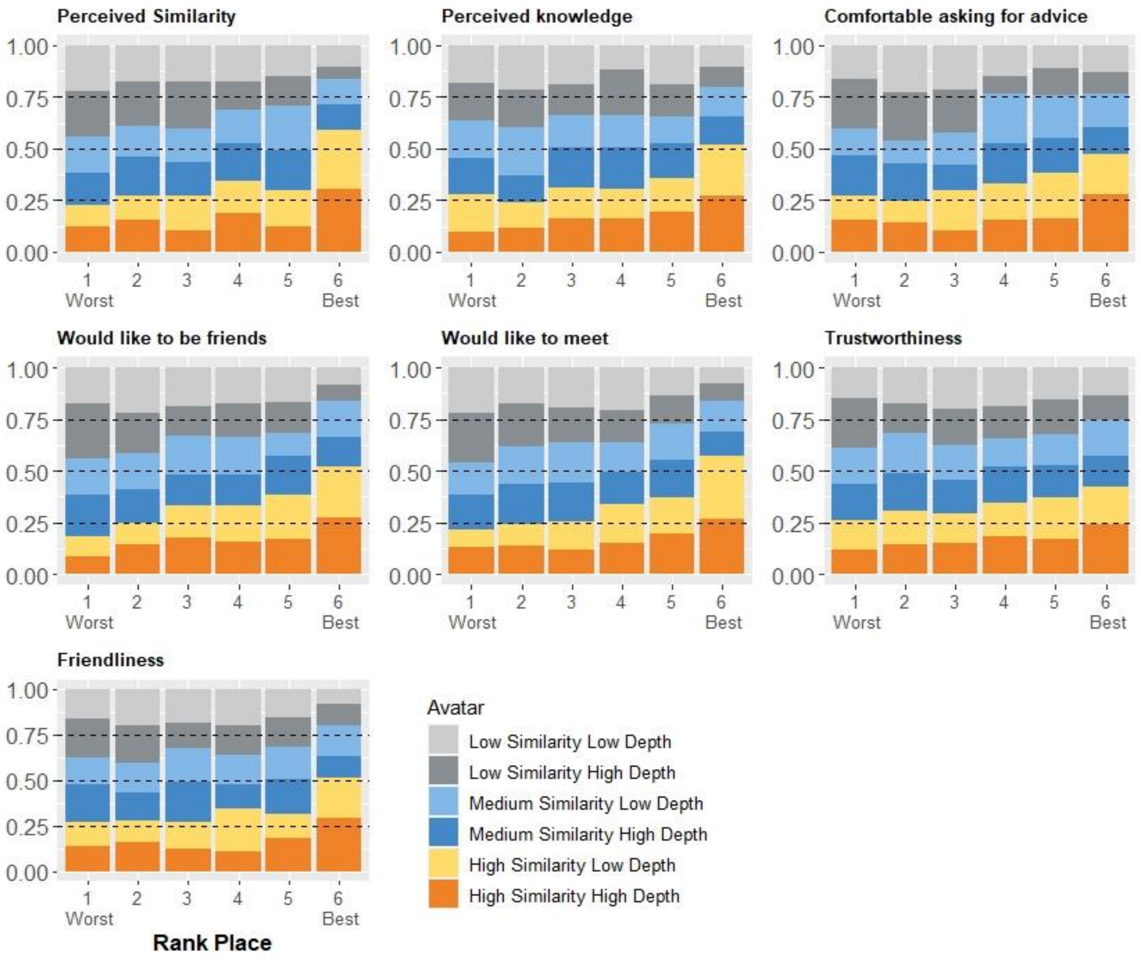


Figure 2-2 Stacked bar plots for the ranking items in Experiment 1. X-axis: rank place from 1 ("Describes Worst") to 6 ("Describes Best"). Y-axis: cumulative percentage of participants who placed each avatar in the corresponding ranking categories.

2.2.4 Discussion

In this experiment, we found a highly consistent and significant effect of similarity on all dependent measures and no effect of disclosure depth or any interaction between the two factors. Participants showed greater preferences for avatars whose self-disclosures indicated greater similarity to themselves, but not for those whose self-disclosures were of greater depth. Interestingly, participants not only perceived the high-similarity avatars as more similar to themselves, but also as the ones that they had gotten to know better, even after controlling for the actual number of statements that they had read about the avatars. This might suggest that similarity contributes to a sense of familiarity (Moreland & Zajonc, 1982), which in return contributes at least partially to the increased liking of more similar others (Reis et al., 2011).

2.3 Experiment 2

Experiment 2 attempts to increase the robustness of our disclosure depth manipulation by emphasizing the real social consequences of participants' choices. To achieve this, participants were invited to the lab in groups of seven to promote the deception that they would be working together. As before, participants began the task by getting to know each other on the computer. Here, however, they believed that they would be allowed to interact with some of the people they met in the task at the end of the study session. Specifically, they were told that the computer would match them with some other participants for face-to-face interactions based on their responses on the ranking measures. In reality, there was no interaction. We expected that this set-up would induce stronger effects and potentially enhance the effect of disclosure depth, which was not significant in the previous study. Other than this change, there were no changes to the study methods. The data analysis and report strategies were the same as used in Experiment 1.

2.3.1 Participants

A total of 103 participants completed this experiment. Eleven participants were excluded for being inattentive (N=3, based on the same criterion as used in the previous experiment) or expressing suspicion about the deception during debriefing (N=8). The remaining 92 participants (33 male, 58 female, 1 not specified) were aged 17 to 27 (*M*=18.40, *SD*=1.27) and consisted of 39 Asian (41.9%), 35 Caucasian (37.6%), 6

Middle Eastern (6.5%), 2 Hispanic (2.2%), 5 mixed or other ethnicities (5.4%), and the rest 5 (5.4%) did not provide this information.

2.3.2 Results

Results from this sample replicated findings from Experiment 1. Specifically, we found a significant effect of similarity across all the ranking items but there were no effects of disclosure depth or any interaction. Similarity significantly predicted participants' ranking responses on both perceived similarity, $\chi^2(2)=41.00$, $p<.001$, and perceived knowledge of the avatars, $\chi^2(2)=10.87$, $p=.004$. As in the previous experiment, participants were more likely to rank the high-similarity avatars, $b=1.031$, $SE=0.118$, $p<.001$, and less likely to rank the low similarity higher on perceived similarity, $b=-0.721$, $SE=0.112$, $p<.001$. Participants were also more likely to rank the high-similarity avatars, $b=0.601$, $SE=0.115$, $p<.001$, and less likely to rank the low-similarity avatars as ones that they had gotten to know better, $b=-0.373$, $SE=0.109$, $p<.001$. Figure 2-3 shows the proportions of participants that placed each avatar in each ranking position for each of the measures.

As above, similarity also had a significant effect on participants' rankings on how comfortable they would feel asking the avatars for advice, $\chi^2(2)=23.14$, $p<.001$, how much they'd like to admit the avatars to their circle of friends, $\chi^2(2)=33.09$, $p<.001$, and how much they'd like to actually meet them $\chi^2(2)=21.70$, $p<.001$. Participants were more likely to rank the high-similarity avatars as the ones from whom they would feel more comfortable asking advice, $b=0.703$, $SE=0.114$, $p<.001$, whom they would most like to admit to their circle of friends, $b=0.819$, $SE=0.115$, $p<.001$, and whom they would more like to meet in person, $b=0.773$, $SE=0.117$, $p<.001$. Low-similarity avatars, on the contrary, were less likely to be ranked higher on any of these three items ($b=-0.516$, $SE=0.111$, $p<.001$; $b=-0.647$, $SE=0.112$, $p<.001$; $b=-0.568$, $SE=0.111$, $p<.001$, respectively in the order of questions).

Finally, we observed a significant effect of similarity on participants' ranking of avatars' perceived trustworthiness, $\chi^2(2)=22.03$, $p<.001$, and friendliness, $\chi^2(2)=15.086$, $p<.001$. Participants were more likely to rank the high-similarity avatars higher on both trustworthiness, $b=0.456$, $SE=0.111$, $p<.001$, and friendliness, $b=0.564$, $SE=0.112$, $p<.001$. The low-similarity avatars were less likely to be ranked higher on either trustworthiness, $b=-0.499$, $SE=0.110$, $p<.001$, or friendliness, $b=-0.491$, $SE=0.111$, $p<.001$.

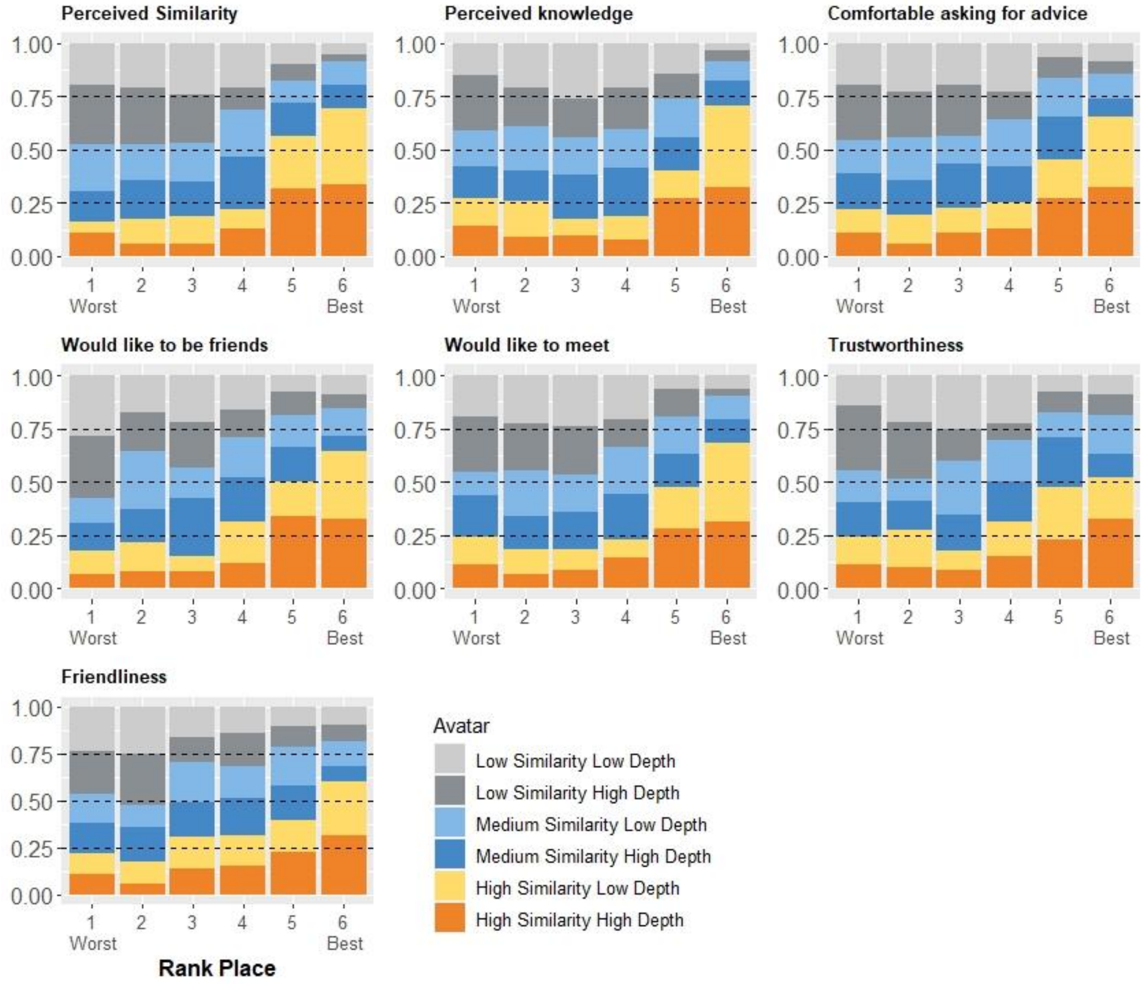


Figure 2-3 Stacked bar plots for ranking items in Experiment 2. X-axis: rank place from 1 ("Describes Worst") to 6 ("Describes Best"). Y-axis: cumulative percentage of participants who placed each avatar in the corresponding ranking categories.

2.3.3 Discussion

Experiment 2's results replicated our previous findings. Higher similarity was associated with greater preferences and more positive impressions across all the ranking items, whereas we found no effects of disclosure depth or an interaction between these factors. Interestingly, the Experiment 2 effects were somewhat larger in magnitude compared to the ones in Experiment 1. This suggested that, as expected, engaging participants by promoting the belief that they were actually interacting with other people did lead to a more robust manipulation than in the previous impression formation study. Regardless, the anticipated effect of disclosure depth failed to materialize. Together, these findings suggest that although similarity seems to be an important element in self-disclosure, the depth of a disclosure alone is not a critical element of disclosure outcomes.

However, one aspect of self-disclosure that may be important is its signaling function. That is, a higher-depth disclosure may signal liking, trust or another prosocial intention on the part of the sender. In this experiment, participants knew that the computer had randomly selected the disclosures from a list of possible items. Thus, even though the statements that participants read were first-person statements (e.g., “I am always procrastinating and would like to change that.”), participants might not have perceived any *intention* to disclose this information. This might be especially relevant to our disclosure depth manipulation. For example, previous research has suggested that people might like deeper disclosures more because such disclosures communicate greater liking and affiliative intentions (Taylor, 1979). Our next experiment addresses this idea.

2.4 Experiment 3

In this experiment, we made the explicit selection of disclosures a feature of the experimental design. This allowed us to ask whether the intention to disclose information makes disclosure depth an important feature of the disclosure process. The study design was exactly the same as in Experiment 2, with one critical modification. Specifically, just before the disclosure phase, participants viewed five randomly selected statements from each “fellow participant”. After reading these statements, participants chose 15 of their own statements to share with that avatar. Consequently, they believed that most of the statements that they saw in the information seeking stage of the session were intentionally selected for them by the other participants. To examine whether participants would indeed perceive the high-disclosure-depth avatars as expressing greater affiliative intentions, we added one item to the ranking measures. Participants were asked to rank the avatars based on “How much you think they’d like to be your friend” in addition to the other ranking questions. The data analysis and report strategies were the same as used in Experiment 1 and 2.

2.4.1 Participants

A total of 95 university students took part in Experiment 3. We excluded 13 participants for being inattentive (N=5, based on the same criterion as used in the previous experiment) or expressing suspicions about the manipulation during debriefing (N=8), as in the previous studies. The remaining 82 participants (38 male and 44 female) were aged 18 to 38 (*M*=19.11, *SD*=2.51) and consisted of 33 Caucasian (40.24%), 27 Asian (32.92%), 7 Middle-Eastern (8.53%), 5 Hispanic (6.10%), 6 Mixed or Other Ethnicity (7.31%), and 4 people (4.88%) who did not respond to this item.

2.4.2 Results

These results replicated the findings in our previous studies. We found a significant similarity effect across all of our ranking measures but no effects of disclosure depth or any interactions. Similarity significantly predicted participants’ ranking responses for both their perceived similarity, $\chi^2(2)=29.81$, $p<.001$, and their perceived knowledge of the avatars, $\chi^2(2)=7.22$, $p=.027$. Participants were more likely to rank the high-similarity avatars, $b=0.819$, $SE=0.120$, $p<.001$, and less likely to rank the low-similarity avatars as being more similar to themselves, $b=-0.602$, $SE=0.117$, $p<.001$. They were also more likely to rank the high-similarity avatars, $b=0.542$, $SE=0.119$, $p<.001$, and less likely to rank the low-similarity avatars as the ones that they had gotten to know better, $b=-0.394$, $SE=0.114$, *p*<.00,. See Figure 2-4 for the proportion of participants who placed each avatar at each rank place for the ranking items.

For the preferences items, we again found that similarity significantly predicted participants’ rankings on how comfortable they’d feel asking the avatars for advice, $\chi^2(2)=18.39$, $p<.001$, how much they’d like to admit the avatars to their circle of friends,

$\chi^2(2)=22.89$, $p<.001$, and how much they'd like to meet the avatars, $\chi^2(2)=17.29$, $p<.001$. The high-similarity avatars were more likely to be ranked highly on all these three items ($b=0.569$, $SE=0.117$, $p<.001$; $b=0.601$, $SE=0.117$, $p<.001$; and $b=0.613$, $SE=0.118$, $p<.001$, respectively in the order of the questions), and the low-similarity avatars were less likely to be ranked higher on either of the three items ($b=-0.451$, $SE=0.115$, $p<.001$; $b=-0.426$, $SE=0.115$, $p<.001$; and $b=-0.400$, $SE=0.115$, $p<.001$, respectively in the order of the questions).

Similarity marginally predicted participants' rankings of how trustworthy the avatars were, $\chi^2(2)=5.926$, $p=.052$, and significantly predicted their rankings of how friendly the avatars were, $\chi^2(2)=7.621$, $p=.022$. The high-similarity avatars were more likely to be ranked higher on both perceived trustworthiness, $b=0.260$, $SE=0.113$, $p=.021$, and perceived friendliness, $b=0.329$, $SE=0.115$, $p=.004$. Conversely, the low-similarity avatars were less likely to be ranked higher on trustworthiness, $b=-0.337$, $SE=0.115$, $p=.003$, or friendliness, $b=-0.391$, $SE=0.115$, $p<.001$.

Interestingly, participants did not perceive the high-disclosure-depth avatars as showing greater affiliative intentions than the low-disclosure-depth avatars as we expected. Instead, there was a significant similarity effect on participants' rating of each avatars' friendship intentions, $\chi^2(2)=27.34$, $p<.001$. Again, participants were more likely to rank the high-similarity avatars, $b=0.599$, $SE=0.118$, $p<.001$, and less likely to rank the low-similarity avatars, $b=-0.511$, $SE=0.116$, $p<.001$, as the ones that would more like to be their friends. Thus, self-disclosed similarity appears to signal affiliation intention to a greater degree than does disclosure depth.

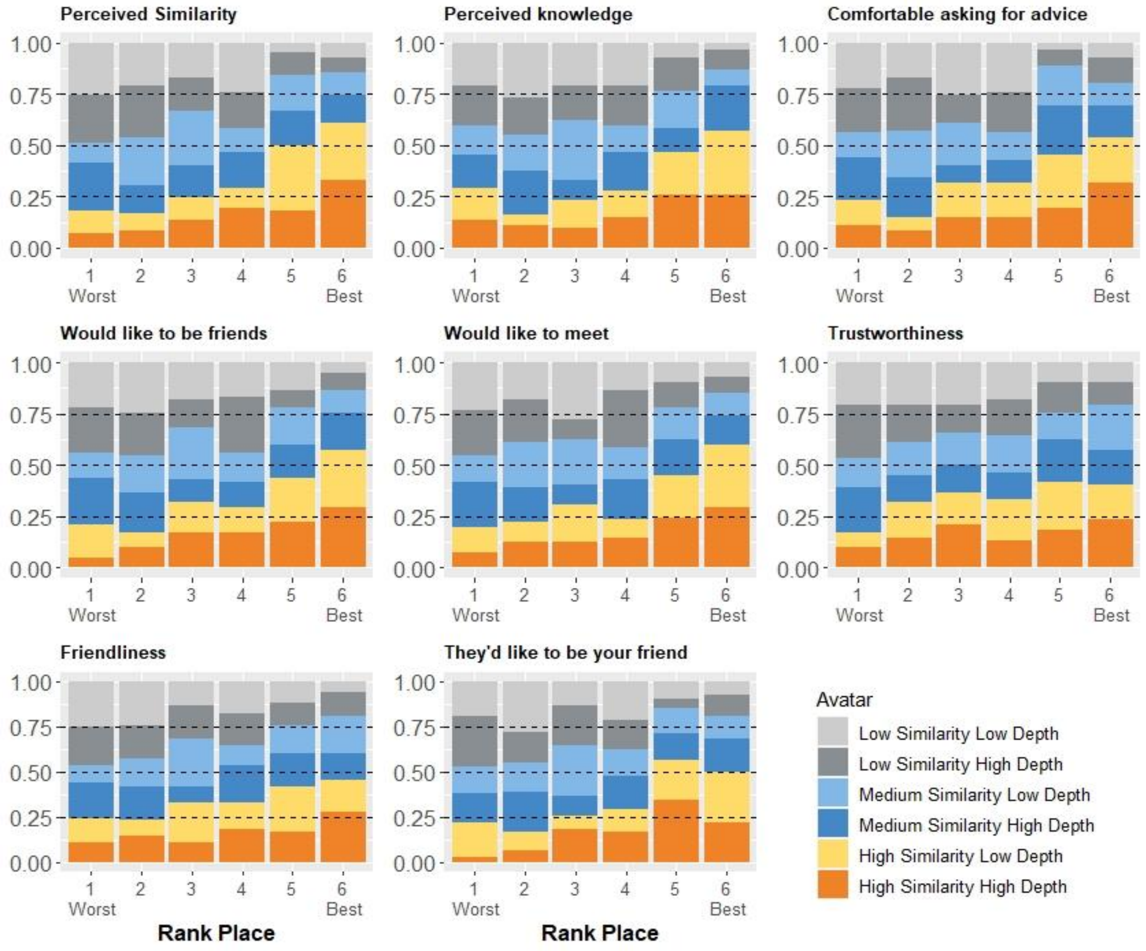


Figure 2-4 Stacked bar plots for ranking items in Experiment 3. X-axis: rank place from 1 ("Describes Worst") to 6 ("Describes Best"). Y-axis: cumulative percentage of participants who placed each avatar in the corresponding ranking categories.

2.4.3 Discussion

As above, we continued to find significant similarity effects and no effect of disclosure depth or any interaction between these factors across our measures. Specifically, participants ranked the high-similarity avatars more positively, and low-similarity avatars less positively on all measures, compared to an "average" avatar. In addition, participants perceived stronger friendship intentions from avatars that self-disclosed greater similarity rather than those who chose to self-disclose more intimately to themselves.

Once again, these findings seem to suggest that similarity, rather than disclosure depth, accounts for the effects of self-disclosure on liking in the previous literature (e.g., Collins & Miller, 1994). Nonetheless, we would like to consider two alternative explanations for our consistently null effect of self-disclosure depth. First, we selected the high disclosure-depth questions and low disclosure-depth questions based on ratings from the 139 participants recruited in a preparatory phase. One explanation for our failure to uncover disclosure depth as a factor in our results may be that the specific participant samples used in Experiments 1 to 3 did not perceive differences in the depth of our high- and low-depth statements. A second issue may be that despite the advantages of the ranking measures (e.g., reducing inattentive or careless responding; the ability to examine relative differences between avatars), it was possible that this forced ordering format might have biased participants’ responses. Specifically, participants believed that their ranking responses would be used to decide whom they would meet in later in-person interactions. They therefore might have put more thought into placing their most and least favorite avatars in the corresponding rank positions and less thought into placing the middle-ranged avatars (McCarty & Shrum, 2000). This might have inadvertently led to an overweighed influence of the similarity effect and obscured the more subtle effect of disclosure depth. We further addressed these two potential issues in the next two experiments.

2.5 Experiment 4

Experiment 4 used the same procedure as in Experiment 3 to ensure that participants believed that the statements that they viewed during the information seeking stage were intentionally selected for them by the other participants. To address potential biases of the ranking measures, we converted the ranking items into 7-point Likert scale items. As in Experiment 3, we again asked participants to indicate how much they thought the avatars would like to be their friend. For each of these questions, participants rated each avatar from 1 (“Strongly Disagree”) to 7 (“Strongly Agree”). We retained one ranking item “How much you’d like to meet them” to allow direct comparison of the results with the previous studies. The number of clicks on each avatar were recorded as in the previous studies. Additionally, to verify that our high and low disclosure-depth questions were indeed perceived as such by the participants, they also rated each of the 40 multiple-choice questions on the degree to which they perceived these items to be personal on a scale from 1 (“Not at all personal”) to 7 (“Extremely personal”).

In addition to addressing these potential methodological issues, we implemented one extra manipulation in this experiment. The literature suggests people might like a discloser even more when they make a *personalistic attribution* of the self-disclosure (Collins & Miller, 1994). That is, when a receiver makes an internal attribution about others’ self-disclosing behaviour and interprets the disclosure to the special qualities that they themselves possess (e.g., likeable, trustworthy, understanding, etc.), they might respond especially positively to the discloser. It is possible that a high level of personalistic attribution would facilitate the effect of deep self-disclosures on liking because people assume a stronger friendship intention from someone who selectively discloses personal information to them and not to others.

To test this idea, we retained the similarity and disclosure-depth manipulations from Experiment 3 but incorporated an additional manipulation. Specifically, participants were told which pieces of information each avatar had shared with them alone (i.e., highly personalistic disclosures), versus with “other participants” as well (i.e., low-personalistic disclosures). Due to a concern that increasing the number of avatars would reduce the believability of the manipulation and induce additional difficulty for participants in remembering information, we applied this manipulation on a between-subjects basis. The computer randomly assigned half of the participants to the high personalistic condition, and the rest the low personalistic condition using a double-blind design to minimize the potential for experimenter effects.

To achieve this manipulation, participants were told that the statements that they would view in the disclosure phase were colour-coded such that statements appearing in one colour had been made to them and “no other participant”. Statements appearing in another colour were those that had been shared with “at least one other participant”. For participants in the high personalistic condition, 50% of the statements that they read from each avatar were coloured as having been made only to them and no one else. In contrast, for those in the low personalistic condition only 25% of the statements they viewed were coloured as having been made only made to them.

2.5.1 Participants

A total of 166 participants took part in Experiment 4. In addition to the exclusion criteria used in the previous experiment, we also excluded “inattentive” participants who spent less than 250 milliseconds on each of 10 or more of the rating questions. We thus excluded a total of 28 participants for expressing suspicions during debriefing (N=8), clicking on the avatar pictures too few times (N=7), or being inattentive based on the reaction time criterion just described (N=13). The remaining 140 participants (44 male, 96 female) aged from 17 to 30 (*M*=18.60, *SD*=1.54) consisted of 69 Asian (48.3%), 45 Caucasian (32.1%), 8 Middle Eastern (5.7%), 2 people of African descent (1.5%), 2 of Hispanic descent (1.5%), 7 mixed or other ethnicity (5.0%), and 7 (5.0%) who did not report this information.

2.5.2 Hypotheses and data analysis

We hypothesized main effects of similarity, disclosure depth, and personalistic attribution, as well as a similarity * disclosure depth interaction and a disclosure depth * personalistic attribution interaction on participants social preferences and impressions of the avatars. These terms were used as the predictors in all our models. Again, the number of clicks was also included in the models to control for any confounding effects of knowledge on liking.

For the rating data, we aggregated participants’ ratings across the social preference items and the perceived traits items, respectively. We then fit linear-mixed models, with the aforementioned predictors, to the aggregated ratings using the *lmer* function in the *lme4* package (Bates et al., 2015) in R (version 3.6.3). We used linear-mixed models to capture the hierarchical nature of the data: as ratings of the avatars were nested within each participant, we allowed the intercept to vary across subject. Again, a global sum-to-zero contrast was set to compare different manipulation levels to the grand mean. Here we first report the Type III Analysis of Variances (ANOVA) results for any significant overall effects of similarity, disclosure depth, personalistic attribution, or either of the interaction terms. The coefficients associated with any significant predictors are then reported to show the magnitude and direction of the effects.

For the ranking responses, the same data analysis and report strategies were used as in the previous experiments. Full details of all models and their estimated effects in Appendix B.

2.5.3 Results

2.5.3.1 Disclosure depth verification

We calculated participants’ ratings of the degree to which the items constituted personal information across the 20 low-disclosure-depth items and the 20 high-disclosure-depth items, respectively, for each participant. A paired-samples t-test found that participants indeed considered the high-disclosure-depth items (*M*=3.20, *SD*=0.97) as significantly more personal than the low-disclosure-depth items (*M*=1.74, *SD*=0.70), *t*(139)= 23.44, *p*<.001. This suggests that our manipulation of disclosure depth had not fail simply due to idiosyncratic differences across participant samples or a failure to select reasonable high-depth items.

2.5.3.2 Rating measures

We failed to find any significant effects of similarity, disclosure depth, or personalistic disclosure on most of our rating measures. No significant effects were found on perceived similarity or perceived knowledge[4]. We calculated the mean ratings across the preference items and the perceived friendship intention item as the aggregated social preference scores[5]. We aggregated the trait perception items by calculating the mean ratings across the two trait perception items. Mixed-linear models with the previously described predictors were fit to the data. We failed to find any effects of the predictors on the aggregated preference scores. We found a significant effect of similarity on the aggregated trait perception scores, $F(2, 701)=3.87$, $p=.021$, in which high-similarity-avatars were associated with lower ratings on this item, $b=-0.073$, $SE=0.035$, $t(701)=-2.10$, $p=.036$. No significant effects of disclosure depth, personalistic attribution, or any interactions were found on this item. See Appendix B for full statistical details.

2.5.3.3 Ranking measure

We included a ranking measure for the question “How much you’d like to actually meet them” to allow for direct comparison with the previous studies. However, due to a technical failure, only 79 participants completed the ranking item. Readers must therefore be cautious of the following results, as were obtained with data from these 79 participants instead of the full sample.

Here, we found a significant effect of similarity, $\chi^2(2)=6.62$, $p=.037$, qualified by a marginally significant interaction between similarity and disclosure depth, $\chi^2(2)=5.15$,

[4] Responses on these two items were not aggregated values and thus analyzed individually as ordinal data with cumulative link models using the *ordinal* R package.

[5] Due to a technical failure, ratings on the “How much you’d like to actually meet them” item were not measured for most of the sample. We therefore removed this item when calculating the aggregated social preference scores.

p=.076. In particular, the high-similarity avatars were more likely to be ranked higher on this item, *b*=0.627, *SE*=0.121, *p*<.001, and the low-similarity avatars less likely to be so ranked, *b*=-0.434, *SE*=0.116, *p*<.001. However, the effect of the low similarity was mitigated in the high disclosure-depth condition, *b*=0.248, *SE*=0.114, *p*=.031. Specifically, participants' decreased interest in meeting the low-similarity avatars was stronger among the low rather than the high-disclosure-depth avatars. See Appendix B for the estimates of all the predictors included in the model. Figure 2-5 shows the distribution of participants responses on the ranking measure.

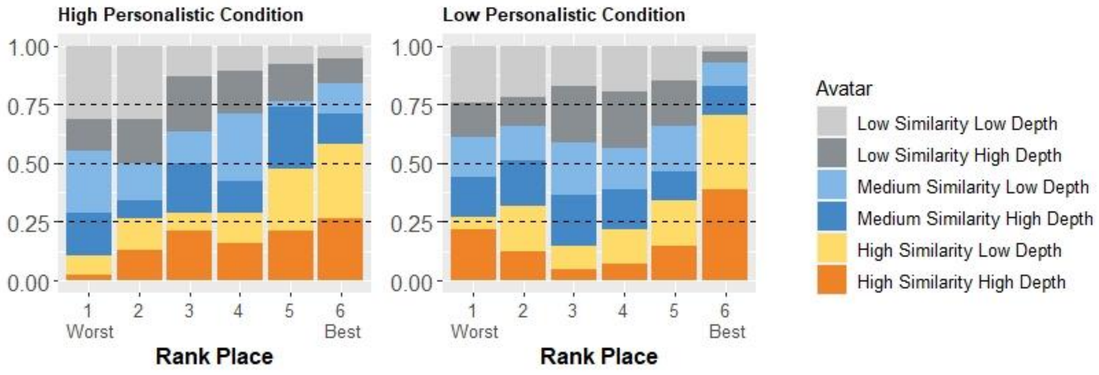


Figure 2-5 Stacked bar plots for the ranking question "How much would you like to actually meet them?" for participants in the high personalistic condition (left) and participants in the low personalistic condition (right) in Experiment 4. Y-axis: cumulative percentage of participants who placed each avatar in the corresponding ranking categories.

2.5.4 Discussion

In this experiment, which used rating, rather than ranking measures, we only found a significant similarity effect on the aggregated trait perception items, but the direction of the similarity effect was inconsistent with any of the previous findings. We found no other significant effects of similarity, disclosure depth, personalistic attribution, or their interactions on any of the other rating measures. We are therefore hesitant to draw any conclusions based on the surprising negative effect of similarity found with the trait perceptions, considering that it was not found on any other rating measures, nor was it consistent with the cumulative evidence in the previous studies.

One likely possibility for the largely null effects found in the rating measures might be that participants were overwhelmed with the amount of information available. The statements that they read in the information seeking stage not only varied in their levels of similarity and disclosure depth, but also were coded in three different colors (i.e., the ones were made only to them; made to them and “other participants”; and the ones randomly selected by the computer). This might have made it difficult for the participants to form clear overall impressions of the avatars as they were able to do in the previous studies. This was partly reflected in the fact that they failed to perceive the different levels of similarity across the avatars.

Interestingly, participants’ ranking responses on the item “how much you’d like to meet them” showed both a significant effect of similarity, as found in the previous studies, and a marginally significant similarity x disclosure-depth interaction effect. Specifically, participants seemed to show more interest in meeting the high-similarity avatars and less interest in meeting the low-similarity-avatars. Moreover, the unfavorable responses to the low-similarity-avatars were observed especially among the low-disclosure-depth avatars, rather than the high-disclosure-depth avatars. This interaction suggests the possibility that higher levels of disclosure depth, though not an independent influence on social preferences, may act as a “buffer” to the negative effects of low similarity between two interaction partners. However, as only a subgroup of participants completed this measure, results should be interpreted with caution.

Finally, participants did rate the 20 high-disclosure-depth questions as being higher in “personalness” than the 20 low-disclosure-depth questions. This suggests that participants did indeed experience the disclosures as differing in depth.

2.6 Experiment 5

In real-world social interactions, self-disclosure is the process by which two people mutually exchange information about themselves in a dynamic interaction. Although the previous studies offered tightly controlled experimental manipulations of similarity and disclosure depth, none of these experimental paradigms allowed for this mutual exchange of information. Thus, it is possible that it is the mutual deepening of disclosure depth, rather than a disclosure’s absolute depth that may be important to forming positive first impressions and developing interpersonal liking after controlling similarity.

Given that the self-disclosures are often exchanged in the development of interpersonal relationships (Altman & Taylor, 1973), it is reasonable to expect that the norm of reciprocity (Cropanzano & Mitchell, 2005; Gouldner, 1960; Laursen & Hartup, 2002) also applies to self-disclosures. That is, self-disclosure depth may matter not because of the information value that it carries, but because it can be exchanged, coordinated, and reciprocated in a social exchange process. Thus, it might not be the depth of the self-disclosures per se, but the reciprocity of disclosure depth that leads to greater liking of a discloser. In our previous experiments, participants simply viewed their social partners’ self-disclosures, excluding this element of reciprocity. In this study, we modified our experimental paradigm to allow participants to engage in back-and-forth disclosure exchanges with the avatars and manipulated the avatars’ level of reciprocity for participants’ own disclosure depth.

2.6.1 Participants

A total of 109 participants completed the experiment. We excluded 5 participants from the analysis: 2 for expressing suspicion during debriefing, 2 for being inattentive (i.e., spent fewer than 250 milliseconds on each of 10 or more rating questions, as in Experiment 4), and 1 for being both inattentive and suspicious. The remaining 104 participants (30 male, 74 female) were aged 17 to 24 (*M*=18.13, *SD*=0.95) and consisted of 54 Asian (51.9%), 35 Caucasian (33.7%), 6 Middle Eastern (5.8%), 3 people of African descent (2.9%), 2 Hispanic (1.9%), 2 mixed ethnicity (1.9%), and 2 who did not provide this information (1.9%).

2.6.2 Methods

As in the previous studies, participants were invited to the lab in groups of seven and completed the task in individual lab rooms, believing that they would meet a few of their fellow participants at the end of the task. As above, the task began with the 40 multiple-choice disclosure questions. Participants then engaged in 10 rounds of self-disclosure exchanges with each of the six avatars, one avatar at a time. The six avatars represented six manipulation conditions in which two within-subject factors, similarity (high vs. low) and reciprocity (high vs. medium vs. low) were fully crossed. The similarity manipulation was implemented similarly to the previous studies. Here, however, the high-similarity avatars self-disclosed information that was the same as participants' own 70% of time and the low-similarity avatars self-disclosed information that was the same as participants' own only 30% of the time.

The reciprocity manipulation was implemented by matching the depth of the avatars' and the participants' own disclosures to different degrees. The high-reciprocity avatars reciprocated the depth of participant's disclosure 80% of time. That is, 80% of the time, if the participant disclosed on one of the twenty high depth questions, the avatars would also disclose on a high depth question; if the participant disclosed on a low depth question, the avatars would also disclose on a low depth question. For the medium- and low-reciprocity avatars, the avatars matched the participants' disclosure depth 50% and 20% of the time, respectively. It is worth noting that the avatars were programmed to only reciprocate the depth of the disclosure, but not topics (which were presented in randomized order). That is, they would not reciprocate participants' self-disclosure by revealing their answer to the same question. Overall, the high-, medium-, and low-reciprocity avatars reciprocated participants' self-disclosure depth 8, 5, and 2 times, respectively, throughout the 10 rounds of disclosure exchanges.

The participant started with the first disclosure of each set of disclosure exchanges and thereafter alternated the order of exchange for the remaining nine rounds to make the process seem more natural and believable. When it was the participant's turn to make a disclosure, they were presented with four pseudo-randomly chosen statements from amongst their answers to the 40 multiple-choice questions. Two of these possible statements came from the high-depth questions and two from the low-depth questions and the computer ensured that a statement that a participant had selected for disclosure to a particular avatar was never repeated with that avatar again (although it might be repeated with another avatar). The computer prompted participants to choose one item from the set of four to disclose to the avatar on that turn. The participant then waited for a random time interval, during which the screen displayed a message that the other participant was choosing their own disclosure to the participant. The avatar's disclosure, as restricted by the previously described manipulations, then appeared on the computer screen for the participant.

After participants finished their exchanges with each avatar, they completed the same social preferences, trait perceptions, and perceived friendship intention rating questions as used in Experiment 4 for that avatar. In addition to perceived similarity, participants also rated the avatar for how much they felt the avatar (1) listened to and (2) responded to what they told them, to measure the perceived reciprocity of the avatar. For all these rating questions, participants responded on 7-point Likert scale ranging from 1 ("Strongly Disagree") to 7 ("Strongly Agree"). They completed these individual ratings after each exchange, before continuing to the next avatar. After completing exchanges with all the avatars, participants ranked all the avatars on how much they'd like to meet them in person. Finally, participants again rated each of the 40 multiple-choice questions based on how personal the question was to verify that the high versus low disclosure depth questions were indeed perceived as different in their overall disclosure depths. In addition to the self-report measures, we recorded the number of disclosures that participants made to the avatars that reciprocated the depth of the avatar's last disclosure to them. These data allowed us to explore whether our manipulation influenced how much *participants reciprocated the avatars' disclosures.*

2.6.3 Hypotheses and data analysis

We anticipated that higher similarity would lead to greater perceived similarity and higher reciprocity would lead to greater perceived reciprocity of the avatars. We also expected higher similarity and higher reciprocity to both lead to greater preference for and more positive perceptions of the avatars. In addition to these hypotheses, we explored whether the manipulation conditions affected participants' own reciprocity to the avatars.

To test these hypotheses, we included similarity, reciprocity, and their interaction as the predictors in our models. As in the previous experiment, we fit linear-mixed models to the continuous dependent variables (e.g., aggregated rating responses; number of reciprocal disclosures to the avatars) and fit a cumulative-link model to the ranking and ordinal responses. The same data analysis and report strategies as in the Experiment 4 were used.

2.6.4 Results

2.6.4.1 Disclosure depth verification

As in Experiment 4, we calculated each participant's mean ratings for how personal the questions were for the high disclosure-depth questions and the low disclosure-depth questions, respectively. A paired-samples t-test again revealed that participants considered the 20 high disclosure-depth questions (*M*=3.07, *SD*=1.00) as generally more personal than the low disclosure-depth questions (*M*=1.53, *SD*=0.54), *t*(103)=22.04, *p*<.001.

2.6.4.2 Rating measures

Responses on the perceived similarity item were treated as ordinal data and analyzed using the cumulative link model. Similarity, reciprocity, and their interaction were included as the predictors. A significant effect of similarity was found, $\chi^2(1)=13.49$, *p*<.001. Specifically, participants were more likely to rate the high-similarity avatars as higher on perceived similarity, *b*=0.454, *SE*=0.074, *p*<.001. No effects of reciprocity or the interaction were found. We averaged responses on the two perceived reciprocity items: how much participants felt that the avatar listened to them, and how much they felt that the avatar responded to them. The aggregated scores were fit to a mixed linear model, with similarity, reciprocity, and their interaction as predictors, while allowing the intercept to vary across participants. Both similarity, *F*(1, 520)=6.73, *p*=.010, and reciprocity, *F*(2, 520)=7.06, *p*<.001, significantly predicted participants' averaged scores on these two items. Specifically, participants perceived the high-similarity avatars, *b*=0.103, *SE*=0.040, *p*=.010, and the high reciprocity avatars, *b*=0.212, *SE*=0.056, *p*<.001, both as higher on how much they listened and responded to the participants. No interaction effect was found. Figure 2-6(a) shows these effects.

We averaged the ratings across all the social preferences items and the friendship intention item for each participant to determine their aggregated social preference score. We also averaged across the two perceived traits (friendliness and trustworthiness) to obtain aggregated trait perceptions scores. We fit both these aggregated measures using the same linear mixed model as above. We found a significant effect of similarity on the aggregated social preference ratings, *F*(1,520)= 30.27, *p*<.001, such that participants significantly preferred the high-similarity avatars, *b*=0.171, *SE*=0.032, *p*<.001. Interestingly, even though the overall effect of reciprocity was not significant, the coefficient for the high-reciprocity avatars was significant and suggested higher preferences of the high-reciprocity avatars, *b*=0.087, *SE*=0.044, *p*=.047. For trait perceptions, we found marginally significant effects of similarity, *F*(1,520)= 3.09, *p*=.080, as well as reciprocity, *F*(1,520)= 2.52, *p*=.082. The direction of the effects were aligned with our expectations: high-similarity avatars, *b*=0.038, *SE*=0.022, *p*=.079 and high-reciprocity avatars, *b*=0.060, *SE*=0.031, *p*=.050, were rated more favorably, whereas low-reciprocity avatars were rated less favorably, *b*=-0.59, *SE*=0.031, *p*=.056. Figure 2-6(b) and 2-6(c) show the effects on the aggregated preferences and aggregated trait perceptions.

2.6.4.3 Ranking measure

We fit the cumulative link model to the participants' ranking responses on how much they'd like to actually meet the avatars, which they completed after interacting with all the avatars, with Similarity, Reciprocity, and their interaction entered as predictors. We failed to find any significant effect of Similarity, $\chi^2(1)$=0.94, *p*=.330, Reciprocity, $\chi^2(2)$=0.24, *p*=.889, or their interaction, $\chi^2(2)$=1.21, *p*=.545.

2.6.4.4 Reciprocity to avatars

Finally, we fit the same mixed linear model used on the aggregated measures to the number of reciprocated disclosures that participants made to the avatars. Here, we found a marginally significant effect of reciprocity, *F*(2,520)=2.93, *p*=0.055. Specifically, participants reciprocated the disclosure depth of the low reciprocity avatars significantly less often, *b*= -0.168, *SE*=0.079, *p*=0.034. We found no effects of similarity, *F*(1,520)=0.16, *p*=0.689, or any interaction, *F*(2,520)=0.31, *p*=0.737. Figure 2-6(d) shows these effects.

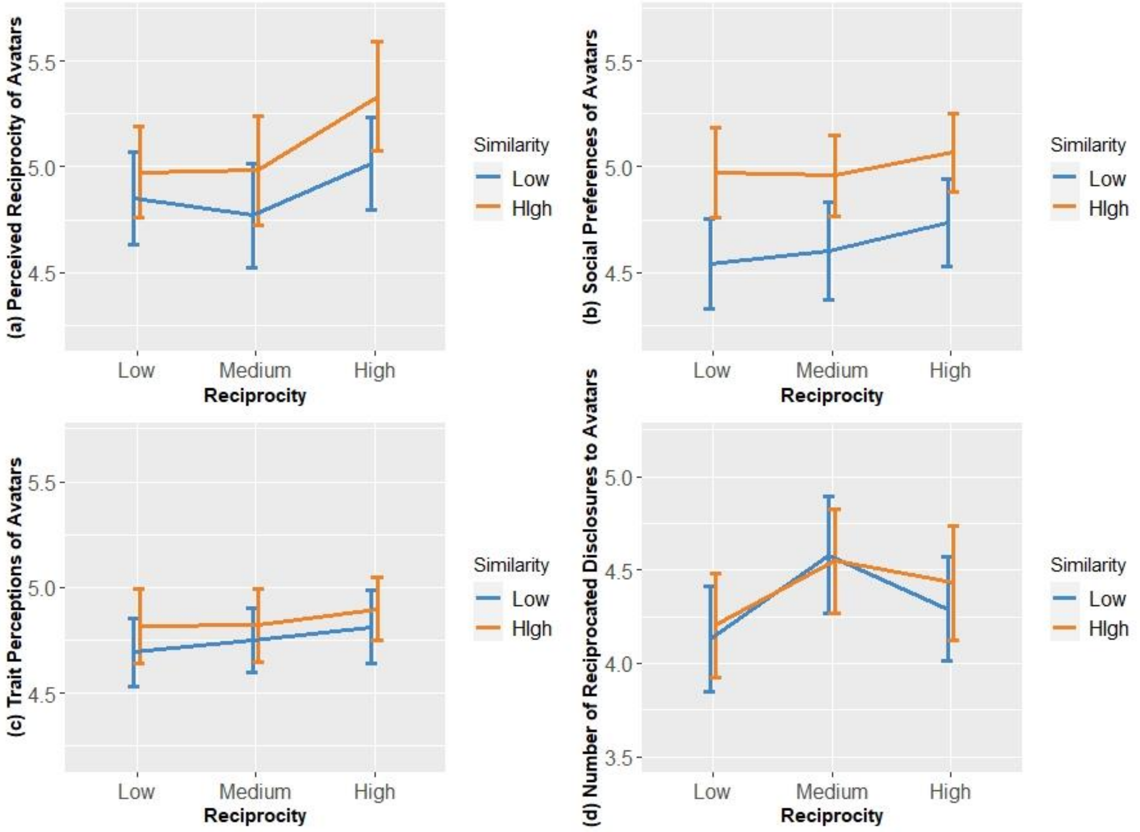


Figure 2-6 The effects of Similarity and Reciprocity manipulations on (a) perceived reciprocity rating, (b)social preference ratings, (c) perceived desirability ratings, and (d) the number of reciprocated self-disclosures to the avatars in Experiment 5. Error bars represent 95% CI.

2.6.5 Discussion

In this experiment, we replicated the significant effect of similarity on participants' impressions and preferences of the avatars as found in Experiments 1-3. Participants rated the high-similarity avatars as higher on perceived similarity, perceived reciprocity, social preference for, and marginally higher on positive perceptions. We found a significant effect of reciprocity on perceived reciprocity of the avatars, such that participants did perceive that the high-reciprocity avatars as being more attentive and responsive. In addition, the effects of reciprocity on participants' preferences and perceptions of the avatars also tended towards statistical significance in the expected direction.

Notably, we used rating measures instead of ranking measures in assessing participants' responses and found similar effects to most of our previous research. Thus, the previous findings were not likely biased by the force-choice format of ranking questions. As in Experiment 4, participants also rated the high disclosure-depth questions as more personal than the low-disclosure depth questions, which increased our confidence in the validity of our disclosure depth manipulation in the previous studies. However, we did not replicate the previously found effect on the ranking item implemented in this study. We suspect that the was partially due to the fact that participants were asked to respond to the ranking item after they completed exchanges with all the avatars, which was an extensive process. Because of the considerable time between the first round of social exchanges and the ranking item, participant might have forgotten the exchanges they had with the earlier avatars. This would help explain why we observed the similarity effect on the rating items, which were completed right after their interactions with each avatar, but not on the ranking item, completed at the end.

Participants rated the avatars who reciprocated their disclosure depth more frequently as both more socially attentive and responsive, indicating that they might have developed more positive impressions of avatars who reciprocated more frequently. The high reciprocity avatars were also associated with higher ratings on preferences and desirable trait perceptions, even though the overall reciprocity effect did not quite reach statistical significance. In addition, participants were also less likely to reciprocate the disclosure depth of the low-reciprocity avatars. These findings suggested that, compared to the previous manipulations of absolute disclosure depth, the manipulation of depth reciprocity might more closely mimic how self-disclosure depth influences interpersonal liking.

2.7 Meta-analysis on the effects of similarity and disclosure depth

To consolidate our findings across experiments regarding the effects of similarity and disclosure depth, we conducted a fixed-effect meta-analysis on the unstandardized logistic regression coefficients for participants' ranking responses on the item "how much you'd like to actually meet them" in Experiments 1 through 4, using the *metaviz* package (Version 0.3.1) in R. We believe that participants' responses on this item most genuinely reflected their impressions and preferences of the avatars, because participants believed that they would be matched with the others for face-to-face interactions based on their responses on this item. Experiment 5 was not included in the meta-analysis because we manipulated the *reciprocity* of disclosure depth rather than disclosure depth itself; the results were thus not directly comparable to findings from Experiments 1 through 4.

As can be seen in Figure 2-7, the meta-analysis suggested that participants consistently showed greater interest in meeting the avatars who appeared to be highly similar to themselves. The summary coefficient for the high-similarity avatars across Experiments 1 to 4 was *b*=0.65 (*95% CI*= 0.55 to 0.75), which corresponds to an odds ratio of 1.92. In other words, participants were about two times more likely to rank the high-similarity avatars higher as the ones they would like to meet in person compared to an "average" avatar across our experiments. In comparison, the summary coefficient for the low-similarity avatars across the four experiments was *b*=-0.48 (*95% CI*= -0.58 to -0.38), corresponding to the odds ratio of 0.62, which suggested that participants were about

40% less likely to rank the low-similarity avatars higher as the ones that they would like to meet in person compared to an “average” avatar. In contrast, participants’ interest in meeting the avatars was not swayed by avatars’ self-disclosure depth, as revealed by the near-zero summary coefficient across the four experiments.

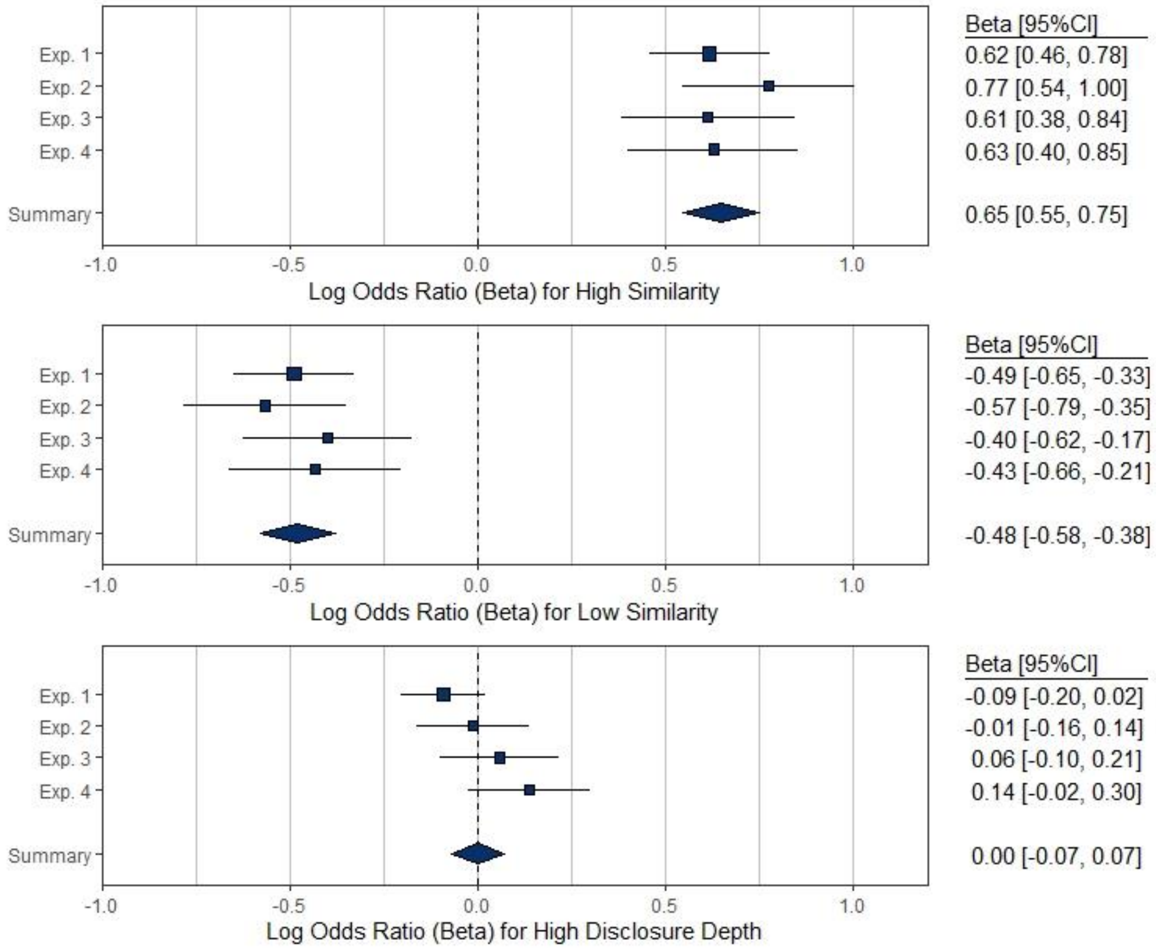


Figure 2-7. Meta-analysis of the effects of High Similarity, Low Similarity, and High Disclosure Depth on participants’ responses on the "how much you’d like to actually meet them" ranking item in Experiments 1 through 4. Positive beta represents higher likelihood to be ranked higher on this item.

2.8 Experiments 1 to 5 general discussion

In a series of experiments, we independently manipulated the disclosure depth and similarity communicated through self-disclosures and examined how these two factors influenced liking for a discloser during an initial social interaction. Overall, findings suggested that, a high level of similarity communicated via self-disclosures leads to more positive impressions and greater social preferences for the discloser; whereas a low level of similarity negatively affects these outcomes. Disclosure depth, however, did not seem to influence people's first impressions and social preferences in these initial social interactions. Specifically, in Experiments 1 to 4, we found consistent and robust effects of similarity on participants' ranking responses of the item "how much you'd like to actually meet them". In Experiment 5, we continued to find that similarity is important in shaping initial liking. However, participants also showed a marginal preference for the avatars who were most likely to reciprocate their own disclosure depth, but the overall effect of reciprocity did not reach statistical significance.

Taken together, we did not find evidence to suggest that absolute disclosure depth enhances liking for a discloser, independent of the similarity between the two interaction partners. In fact, the strong effects of similarity suggest that identifying aspects of similarity with another person might well be one of the main functions of self-disclosure exchange. Thus, it is likely that disclosure similarity underlies the positive relationship between self-disclosure and interpersonal liking found in the literature. Interestingly and contrary to expectations, people did not believe that they got to know the avatars who self-disclosed more deeply to a greater extent. Similarly, they did not perceive stronger friendship intention from the deeply disclosing avatars. Rather, similarity predicted participants' feelings of avatars' friendship intentions. Collins and Miller (1994) suggested that receivers might like those who self-disclose deeply because such disclosures communicate the discloser's liking of the receiver and their desire to enhance the relationship with the receiver. Our findings suggest that these processes are more likely a result of a high perceived similarity, rather than the depth of their disclosures.

Findings from Experiment 5 pointed to the possibility that, compared to the other person’s self-disclosure depth per se, people might more readily perceive another’s *reciprocity* of their own disclosure depth, and may subsequently incorporate such perceptions into their impressions of and interactions with the other person. In other words, interpersonal liking might be less a function of absolute disclosure depth than of the feeling that another is matching one’s own depth. Nonetheless, it is possible that effects of disclosure depth on interpersonal liking during initial social interactions are masked by the strong effect of perceived similarity between interaction partners.

Despite these interesting findings, there are a few limitations to the current studies. First, the self-disclosure statements used in all the studies were generated by university students. Though these topics were considered common topics that university students would talk about when getting to know others for the first time, they are inherently more applicable to this specific population. In addition, even though we verified in Experiments 4 and 5 that participants did perceive the high self-disclosure depth topics as more personal than the low self-disclosure depth topics, the high depth topics were not overly personal (with mean ratings around 3.2 on a 7-point scale). Although in a first interaction, overly personal disclosures are often considered inappropriate and shown to lead to negative social outcomes (e.g., Caltabiano & Smithson, 1983; Chaikin & Derlega, 1974), the relatively narrow range of topics might have limited the robustness of our disclosure depth manipulation. Future studies might want to explore broader topics that vary more in their disclosure depth and to test the effects in other populations.

Second, although the computerized interaction set-up used in all the studies allowed us to manipulate the key variables in a well-controlled manner, it lacked many of the important elements that occur in face-to-face interactions. It is possible that people differ in their non-verbal behaviours such as eye contact, gestures, and facial expressions when disclosing something deeper, in comparison to more superficial information about themselves. Because non-verbal behaviours heavily influence interpersonal liking (Boone & Buck, 2003; McGinley et al., 1978; Scherwitz & Helmreich, 1973), deep disclosure-depth might lead to greater liking, not because of its intimate verbal content but via the non-verbal behaviours that accompany it. Future research should seek to investigate this idea.

Finally, our findings and conclusions pertain to short initial interactions with strangers. We suspect that disclosure depth matters, probably to a great extent, in ongoing relationships where there is a lasting, dynamic exchange of responses between two interaction partners. In such relationships, people not only receive self-disclosures, but also respond to such disclosures and receive feedback about their responses. The development of interpersonal relationships likely relies more on dynamic exchanges of responses rather than solely on the content of one party’s self-disclosures (Reis & Shaver, 1988), as the Experiment 5 results suggest. Deeper self-disclosures might thus lead to greater interpersonal liking in such long-term relationships by facilitating deeper responses from the other people, thereby encouraging future exchanges between interaction partners.

Chapter 3

3 Experiment 6: Risk-Taking, Liking, and Self-Disclosure

In this experiment, I examined the second effect reviewed and supported in Collins and Miller (1994): the finding that people self-disclose more to those whom they like more. In addition, I examined how individual differences in risk taking play a role in this process. Using a paradigm modified from the one used in Experiments 1 to 5, I manipulated participants' liking of three avatars, whom they believed were other participants interacting with them in real-time via computer. Participants were then given a chance to engage in self-disclosure to these avatars and measured on their risk-taking tendencies. Participants' perceptions of the avatars and risk-taking tendencies were subsequently used to predict their self-disclosures to the avatars, after controlling variables such as their perceived similarity and demographic characteristics such as gender and age.

3.1 Rationale

3.1.1 Self-disclosure as a risk-taking behaviour

Self-disclosure can essentially be viewed in a risk-taking framework when one considers the parallels between the mechanisms underlying self-disclosure decisions and those underlying risk-taking decisions. First, when self-disclosing one's beliefs, thoughts, and feelings to another person, there are both potential benefits, such as approval and affiliation, and potential costs, such as interpersonal rejection, associated with personal revelations. It is therefore reasonable to expect people to weigh the potential benefits against costs when they are making decisions regarding whether and what to self-disclose to a particular receiver. Indeed, Taylor and Altman (1975) found that, when randomly assigning participants to interact with a confederate, the amount of time that the sailors spent self-disclosing to the confederate varied as a function of the confederate's responses to them. Those who interacted with a disagreeable and disapproving confederate (i.e., higher costs of disclosing) spent significantly less time self-disclosing compared to those who interacted with an agreeable and approving confederate (i.e. higher rewards of disclosing). Such cost-benefit calculation also lies at the core of decision making involving risks (Bernoulli, 1954/1783; Harless & Camerer, 1994; Tversky & Kahneman, 1973). For example, in his pioneering work on risk measurement, Daniel Bernoulli (1954/1783) suggested the importance of "utility", one's subjective evaluation of an item's value. People's risk-taking decisions are a direct product of item utility and risk. Specifically, one should be willing to take a risk that has a "mean utility" that equals the utility of the item. That is, the balance between the potential costs and rewards of the risk, after taking in account of their respective probability of occurrence, should be the same as the subjective value of the item.

The second parallel between self-disclosure and general risky decision making is that they both reflect a learning process whereby an individual adjusts their behaviours based on the feedback from the environment. For example, in Taylor and Altman (1995), when the confederate changed their responses to be more negative or more positive later in the same session, the researchers observed a corresponding change in participants' self-disclosing behaviour, reflecting participants' sensitivity to the changes in the cost-reward ratio. Similarly, Reis and Shaver (1988) explicitly theorized that a receiver's responses to one's self-disclosures would influence whether the discloser feels that they are understood, validated, and cared for by the receiver, which in turn either facilitates or inhibits further disclosures. This process also underlies people's general decision making under risk. For example, Cook et al. (2005) found that participants who played with the same partner repeatedly in a Prisoner's Dilemma game gradually entrusted more money to their partner, whereas those who played with a random partner in each round did not show such a pattern. Their results suggested that repeatedly interacting with the same partner allowed the participants to learn about the environment (i.e., whether this partner would corporate) and adjust their subsequent risk-taking behaviours accordingly.

The third parallel between self-disclosure and risky decision making is that they are both influenced by affect and fluency. Self-disclosures have been shown to be facilitated by positive mood (Cunningham, 1988; Forgas, 2011) and greater processing fluency (Alter & Oppenheimer, 2009b), which is the ease with which people process information (Alter & Oppenheimer, 2009a). For example, Forgas (2011) found that participants self-disclosed more information and with more intimate details after watching a short comedy film or writing about a positive life episode, compared to those who watched a sad short film or wrote about a negative life episode. In an interesting field study, Alter and Oppenheimer (2009a) examined how processing fluency could influence self-disclosures by comparing anonymous confessions on an online confession site two weeks before and two weeks after the website was reformatted, which made it easier to read and thus increased its processing fluency. A group of independent raters rated each confession based on how embarrassed they would be if asked to disclose that information. Confessions after the website reformatting were found to be more embarrassing than those before, indicating that fluency may facilitate deeper levels of self-disclosure. Similarly, positive affect and greater fluency encourage people’s risk taking. The *risk-as-feelings* hypothesis, for example, was raised to address how affect experienced at the time of a decision can influence or even override cognitive evaluations of its risk (Loewenstein et al., 2001; Slovic et al., 2005). It suggests that positive mood facilitates optimistic judgments and decreases risk perceptions, which might in turn lead to greater risk-taking behaviours, whereas negative affect promotes risk perception and thus discourages risk taking behaviours (Loewenstein et al., 2001).

3.1.2 Individual differences in risk-taking tendency and self-disclosures

People vary in how much they are prone to take risks. This variation in risk-taking tendency can be attributed to individual differences in factors such as gender (Byrnes et al., 1999), age (Deakin et al., 2004), and personality traits such as impulsive sensation seeking, aggression-hostility, and sociability (Zuckerman & Kuhlman, 2000). Taken together with the previous argument that self-disclosing behaviours can be understood in a risk-taking framework, individual differences in risk taking might also be reflected in whether and how much people choose self-disclose to their social partners (i.e., to take a social risk). Specially, people higher in general risk-taking tendency might be more willing to self-disclose than those with a lower risk-taking tendency.

Moreover, individual differences in how much people weight potential rewards versus potential costs associated with self-disclosing behaviours might influence how much they self-disclose to a given partner. The *Behavioural Inhibition System (BIS)/ Behavioural Activation System (BAS)* (Fowles, 1980; J. A. Gray, 1981, 1982) is a theoretical framework thought to represent how individuals differ in the extent to which punishment and reward, respectively, motivate their behaviours. The BIS was conceptualized as the motivational system responsible for inhibiting behaviours that might lead to negative outcomes. It is responsive to threatening stimuli such as punishment, frustrative non-reward, and novelty. A higher sensitivity in the BIS is associated with higher anxiety and greater avoidance behaviours when an individual is faced with possible punishment. The BAS, in contrast, is responsible for promoting approach behaviours and is sensitive to signals of reward and non-punishment. A higher sensitivity in the BAS should thus be associated with greater positive affect and greater goal-pursuit when an individual is presented with incentives. The BIS and the BAS have been conceptualized as independent systems, such that individuals could be high or low on either of these two dimensions respectively (Carver & White, 1994; J. A. Gray, 1981; Torrubia et al., 2001). Empirical studies using self-report measures developed based on the BIS/BAS theoretical framework have linked higher BAS sensitivity with increased impulsivity (Braddock et al., 2011), risky health behaviours such as substance use and unprotected sexual practices (Braddock et al., 2011; O'Connor et al., 2009), and increased risk for gambling (Gaher et al., 2015; Kim & Lee, 2011; O'Connor et al., 2009), whereas higher BIS sensitivity was generally found to link with decreased impulsivity and risk-taking behaviours (e.g. Braddock et al., 2011; O'Connor et al., 2009).

Such individual differences on BIS/BAS sensitivity should theoretically influence participants' motivation to avoid the potential punishment and approach the potential reward associated with self-disclosing behaviour in social interactions. As previously discussed, self-disclosing to a social partner might lead to both potential rewards and costs. A person higher on BIS sensitivity should inhibit their self-disclosing behaviour more as they are more motivated to avoid the potential costs, compared to those lower on BIS sensitivity. In contrast, individuals higher on BAS sensitivity should be more motivated by the potential rewards and thus engage in greater self-disclosing behaviour. Furthermore, these relationships between BIS/BAS sensitivity and self-disclosing behaviour might be exacerbated when the discloser likes the receiver: greater liking for a receiver might make their approval more rewarding and the potential rejection from them more dreadful. It is therefore reasonable to expect those high on BIS sensitivity to further inhibit self-disclosing behaviours when interacting with a liked social partner, whereas the same situation would motivate those high on BAS sensitivity to engage in even greater self-disclosing behaviours, compared to interactions with a less liked partner.

3.2 Methodology

3.2.1 Procedure

Participants were invited to the lab in groups of four and were told that they would be interacting with each other first on the computer, followed by potential face-to-face interactions later in the study. As in the previous studies, participants in fact interacted with pre-programed avatars on the computer and no face-to-face interactions took place in the study. To further incentivize attentive behaviours and engagement in the study tasks, participants were informed that they could earn up to $5 bonus money based on their performance in the tasks. Participants were paid accordingly based on their accumulated points at the end of the study, in addition to any compensation that all participants earned by taking part of the study.

As in the previous studies, participants completed the entire study on a computer in individual lab rooms. Participants first chose an animal picture as their profile picture and completed 20 multiple-choice questions. Half of these questions were of high disclosure-depth and the rest were of low disclosure-depth. These questions were randomly chosen, with the restrictions on the depth, from the 40 questions used in the previous experiments. Participants were then told that, for the first task, the computer would randomly pair them up with two of the three other participants and they would play a joint game with each of these two participants. The game was a modified joint Flanker task (Atmaca et al., 2011), which aimed at eliciting different levels of liking toward the "other participants". Participants were told that their joint performance with each partner would contribute to the number of points they won and thus the amount of bonus cash that they would earn by the end of the study. They completed the task with two avatars. Unbeknownst to the participants, one avatar was programed to perform better and contribute more points to the joint score than the other. The third avatar, with whom participants did not play the game, served as a "neutral" partner for whom liking was not manipulated.

In our joint Flanker task, the participant and their avatar partner were each assigned two letters that were randomly selected from a list (e.g., participant--"H", "K"; partner – "S", "C"). On each trial, participants viewed a string of five stimulus letters for 100ms (e.g., "HHHHH", "HHSHH". After the stimulus disappeared, participants pressed one of two keys to indicate whether the central letter in the string was their own letter or belonged to their partner. If they pressed the correct key, they earned points. If the press was slow (i.e., RT>400ms), they earned 1 point; if they pressed the correct key quickly (i.e., RT≤ 400ms), they earned 2 points; if they pressed the incorrect key, they lost 1 point. The computer provided feedback on both the participant's and their avatar partner's performance (e.g., how many points they each earned) as well as the joint points they earned each trail. The better-performing avatar (i.e., the "good" avatar) reacted faster to its target letters and made fewer mistakes throughout the trials than the worse-performing avatar (i.e., the "bad" avatar). See Figure 3-1 for an illustration of the task. The expectation was that participants would develop a stronger liking for the better-performing avatar, as a result of the better task outcome. To verify the effect of our liking manipulation, participants reported their mood using the PANAS scales (Watson et al., 1988) both at the beginning of the study and right after they completed the joint Flanker task with each of the two avatars.

Next, participants were told that they would get to know more about the other participants by reading some statements that the others chosen on the 20 multiple-choice questions at the beginning of the study. For all the avatars, half of their statements were the same as the participant's own choices. In addition, all avatars' statements consisted of

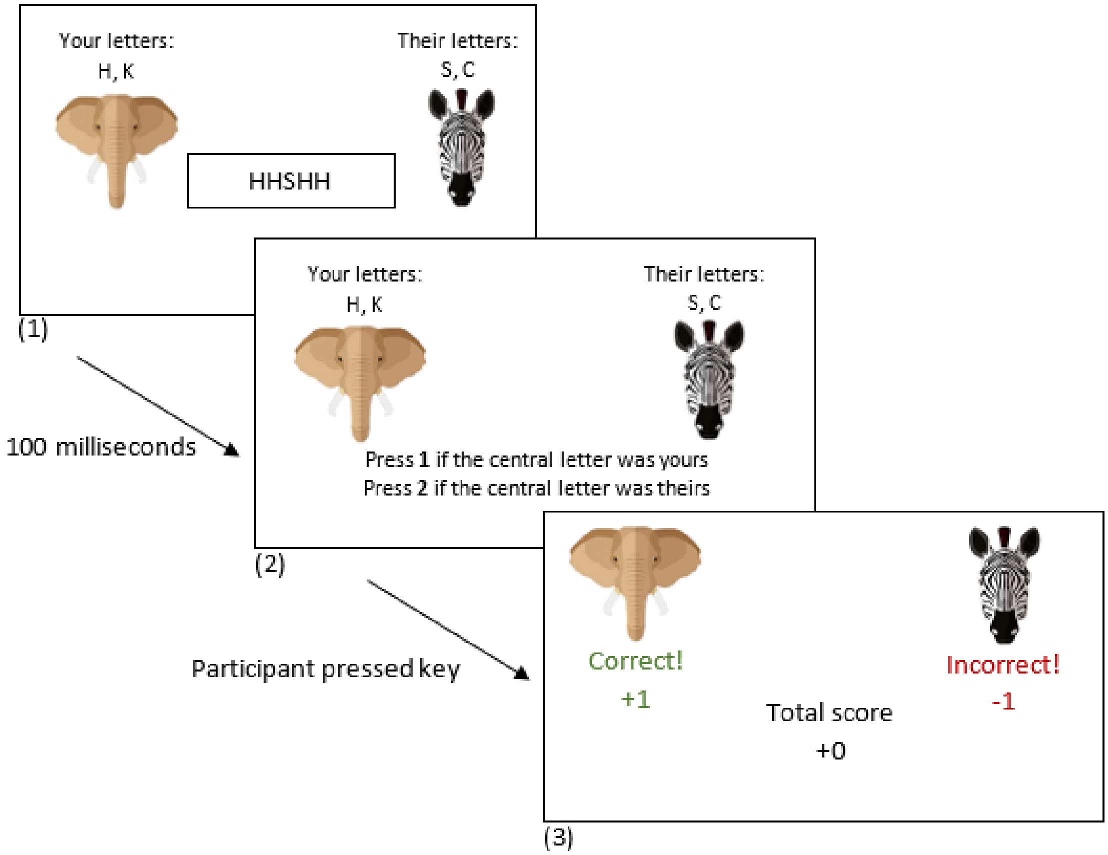


Figure 3-1 On each trial of the joint Flanker Task in Experiment 6, participants were presented with the string at the center of the screen for 100 milliseconds (1). They were them prompted to press 1 or 2 after the string disappeared (2). The computer then displayed the result of the trial (3).

50% low-depth disclosures and 50% high-depth disclosures. This served to equalize avatars actual similarity and disclosure depth, to avoid confounding the effect of our liking manipulation on the dependent variables. After viewing each avatar’s statements, participants completed self-report questions assessing perceived similarity, perceived knowledge, and liking of that avatar.

Upon completing all the self-report liking questions, participants were told that they had a chance to share some more information with each of the other participants. They were told that they would later view each other participant’s self-disclosures to decide whom they’d like to meet in person. They then complete the remaining 20 multiple-choice questions from the list of 40 questions about themselves and chose which ones to share with each avatar. They were also given the chance to freely disclose any additional information that they wanted to share with each avatar. The number of disclosures that they chose to share from the 20 questions, the proportion of deep disclosures they selected, and the number of additional self-disclosures they typed were used as indications of their self-disclosure level to each avatar (i.e., the dependent variables).

Next, participants were told they would complete some measures of personality before viewing the other participants’ self-disclosures. They then completed two behavioural risk-taking measures, the Balloon Analogue Risk Task (BART; Lejuez et al., 2002)) and the Columbia Card Task (CCT; Figner, Mackinlay, Wilkening, & Weber, 2009) and a few self-report questionnaires measuring individual differences in risk-related tendencies. Again, to encourage attentive behaviours and genuine responses, their performances in the BART and CCT were converted into points that they earned at the end of the study. The order in which participants completed the two behavioural risk-taking tasks and the questionnaires were randomized. Lastly, participants completed a funnel debriefing as in the previous studies, paid their bonus earnings, thanked and dismissed.

3.2.2 Hypotheses

Hypothesis 1. Greater liking of an avatar would predict greater self-disclosure to the avatar (i.e., higher number of self-disclosures, higher proportion of deep self-disclosures, and greater number of extra self-disclosures typed in the optional text box).

Hypothesis 2. Greater general risk-taking tendency would predict greater overall self-disclosure across avatars.

Hypothesis 3. Participants' individual BIS/BAS sensitivity would interact with liking of an avatar to predict self-disclosures to the avatar. Specifically, the positive relationship between liking and self-disclosure would be strengthened for individuals with a higher BAS sensitivity and reduced for individuals with a higher BIS sensitivity.

3.2.3 Measures

3.2.3.1 PANAS

The Positive and Negative Affect Schedule (PANAS) is a 20-item self-report scale developed by Watson et al. (1988), which consists of 10 items that assess positive affect and 10 items that assess negative affect. The PANAS has been shown to have good construct validity (e.g., Crawford & Henry, 2004) and test-retest reliability (e.g., Ostir, Smith, Smith, & Ottenbacher, 2005). Participants rated each item in random order on a visual analogue scale (e.g., "interested", "distressed", "excited", etc.) by clicking on a straight line with "very slight or not at all" (0) on the one end and "extremely" (100) on the other end to indicate their current mood. Ratings on the items assessing positive affect and those assessing negative affect were averaged, respectively, to generate participants' Positive Affect and Negative Affect scores.

3.2.3.2 Perceived similarity, knowledge, and liking measure

The same items used in Experiments 1-5 were used to assess participants' perceived similarity, perceived knowledge, social preferences, and trait perceptions for each of the three avatars. Participants rated the avatars on scales ranging from 1 ("not at all") to 7 ("very much so") for each question.

3.2.3.3 Risk taking measures

3.2.3.3.1 The Balloon Analogue Risk Task (BART)

The BART is a laboratory-based behavioural measure that was developed by Lejuez et al. (2002). It has been shown to have good construct validity. For example, higher scores on the BART predicted higher self-reported risk-taking behaviours (Lejuez et al., 2002), alcohol consumption and problems (Fernie et al., 2010), and drug use (Hopko et al., 2006). An fMRI study also showed that when participants played the BART, there was activation in the dopamine rich mesolimbic structures, the brain regions that are consistently activated when risk and reward are involved (Rao et al., 2008).

In this task, participants were presented with a balloon on the computer screen. They could pump air into the balloon by pressing a button on the screen; with each successful pump, participants could earn 1 point. However, each balloon could only take a certain number of pumps, at which point the balloon would "pop" and the participant would lose all the money that they had earned on *this* balloon. Participants experienced 30 trials (i.e., 30 balloons) in the task. The "break point" of each balloon varied and was randomly selected by the computer to range from 1 to 128 pumps, with a mean of 64. The break point for each balloon was unknown to participants. Participants could choose to stop pumping a balloon at any point and collect the money that they had earned on a given balloon by clicking on a "Collect" sign on the screen, which would store their trial earnings to a "permanent bank". A trial immediately ended when the balloon popped or when the participant banked the money, and a new balloon appeared on the screen. The number of times that the participant pumped each balloon was averaged across all the trials, excluding trials in which the balloon popped, and used as an index of general risk taking. See Figure 3-2 for an illustration of the task.

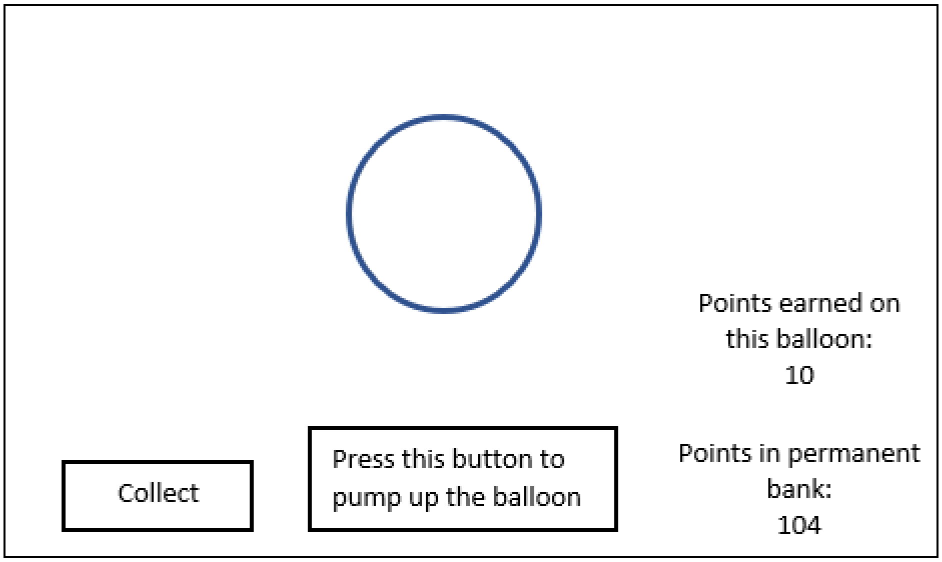

Points earned on this balloon:
10
Collect
Press this button to pump up the balloon
Points in permanent bank:
104

Figure 3-2 The Balloon Analogue Risk Task (BART) used in Experiment 6. Figure adapted from Lejuez et al. (2002).

3.2.3.3.2 The Columbia Card Task (CCT)

The CCT is a more recently developed laboratory-based behavioural measure of risk taking. Figner et al. (2009) developed two versions of CCT—the “hot” version that was designed to assess affective decision making under uncertainty and the “cold” version that was designed to assess more deliberative decision making. Empirical evidence has suggested that the “hot” and “cold” CCT indeed capture different aspects of risk taking (Buelow & Blaine, 2015; Markiewicz & Kubińska, 2015). Interestingly, one study found that, when measured using the “hot” but not the “cold” CCT, those who were highly responsive to rewards showed greater risk taking compared to those with less sensitivity to rewards (Penolazzi et al., 2012). We therefore opted to use the “hot” version of CCT in the current study.

In this task, 32 cards were displayed face-down on the computer screen in each trial. Some of the cards were “gain cards” and some were “loss cards”. A participant could turn over as many cards in a trial as they wished by clicking on them. With each gain card turned over, the participant would gain a certain amount of points for the trial. If a loss card was turned over, the value of the loss on that trial would be subtracted from the participant’s score. Trials ended when either the participant chose to stop turning over cards or when they encountered a loss card.

Participants experienced a total 28 CCT trials. The trials differed in three parameters. First, the gain amount varied. For each trial, the amount of gain per card was either 10 points or 30 points. Second, the probability of loss varied across trials. Either 1 or 3 cards out of the 32 cards were “loss cards” in a trial. Third, the loss amount varied with two possible levels, 250 points or 750 points, across the trials. With two different levels for each of the three parameters, there were 8 possible combinations. The entire task therefore contained 24 trials in which each combination of parameters was presented on 3 trials in a random order. Participants were always informed of these parameters by an information display for each trial. See Figure 3-3 for an illustration of the task.

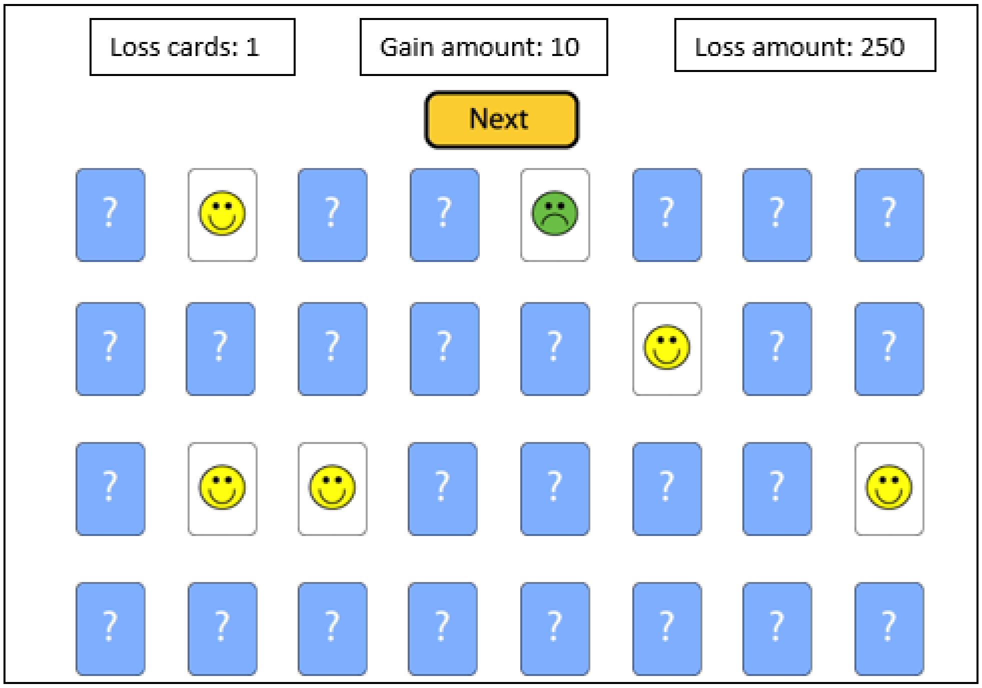


Figure 3-3. An illustration of the Columbia Card Task used in Experiment 6.

As in Figner et al. (2009), we implemented 4 “fake trials” to facilitate believability of the task. These fake trials were programed such that participants would turn over a “loss” card very early in the trial. In comparison, in the other 24 “true trials” previously described, the task was programed such that it was unlikely that participants would turn over a loss card before they voluntarily chose to bank their points and end the trial. This was made possible by programming the trials such that the loss cards were the very last possible cards to be turned over (e.g., in a trial with 3 loss cards, the loss card would not appear until the participant chose to turn over for the 30th time in the same trial). Participants’ risk-taking tendency was calculated by averaging the number of cards that they chose to turn over in the 24 “true trials”.

3.2.3.3.3 The Behavioral Inhibition/Behavioral Activation Scales (BIS/BAS Scales)

Building on the BIS/BAS theoretical framework, Carver and White (1994) developed the BIS/BAS Scale, which is a 20-item self-report questionnaire that assess people's BIS and BAS sensitivity. The BIS/BAS Scale consists of four subscales: the BIS subscale assesses how much one is sensitive to punishment and would experience negative affect when experiencing or anticipating punishment (e.g., "If I think something unpleasant is going to happen I usually get pretty "worked up"."); the BAS Reward Responsiveness subscale captures one's positive affect when obtaining or anticipating reward (e.g., "When I get something I want, I feel excited and energized."); the BAS Drive subscale assesses one's persistent pursuit of a reward or goal (e.g., "When I want something, I usually go all-out to get it."); and the BAS Fun Seeking subscale assesses one's desire for new rewards and willingness to act on the spur of the moment (e.g., "I crave excitement and new sensations"). In the current study, participants rated each question on a scale of 1 ("strongly agree") to 4 ("strongly disagree"). Participants' ratings were subsequently reversed and then averaged across items for each of the four subscales. Higher scores thus represented higher sensitivity of the BIS and the three aspects of BAS.

3.2.3.3.4 The Sensitivity to Punishment and Sensitivity to Reward Questionnaire (SPSRQ)

The SPSRQ is another self-report questionnaire developed by Torrubia et al. (2000) based on Gray's (1981) BIS/BAS theoretical framework. The scale contains 48 yes-no questions, half assessing individuals' sensitivity to punishment (e.g., "Do you often refrain from doing something because you are afraid of being illegal") (i.e., the SP scale) and the rest half assessing their sensitivity to reward (e.g., "Does the good prospect of obtaining money motivate you strongly to do something?") (i.e., the SR scale). Whereas the BIS subscale in the BIS/BAS scales developed by Carver and White (1994) focused on the negative affect associated with anticipated or current punishment, the SP scale captures both the negative affect and behavioural inhibition/avoidance when potential aversive consequences are involved (e.g., “Do you prefer not to ask for something when you are not sure you will obtain it”). The psychometric properties of the scale were supported in other empirical studies (e.g., Beck, Smits, Claes, Vandereycken, & Bijttebier, 2009; O’Connor, Colder, & Hawk, 2004). In the current study, the number of endorsed items on the SP scale were summed up as participants’ sensitivity to punishment scores and the number of endorsed items on the SR scale were summed up as participants sensitivity to reward scores. Higher scores on SP and SR scales thus indicated higher BIS sensitivity and higher BAS sensitivity, respectively.

3.2.3.3.5 The Experience in Close Relationship Scale- Short (ECR-S)

Lastly, we included a measure that specifically captures the approach and avoidance behaviours in social relationships, as it might uniquely predict people’s self-disclosures to others (e.g., Mikulincer & Nachshon, 1991). The Experiences in Close Relationship Scale (ECR) was originally developed in Brennan, Clark, and Shaver (1998), in which the authors presented items from the then-available scales on adult attachment to over 1,000 undergraduate and found two relatively orthogonal dimensions, *Anxiety* and *Avoidance.* Whereas attachment anxiety captures a fear of rejection or abandonment from others, attachment avoidance refers to a fear of dependence and interpersonal intimacy (Mikulincer et al., 2003). Individuals who are low on both attachment anxiety and attachment avoidance are considered as having a secure adult attachment orientation, which manifests as being comfortable with closeness without being overly worried about being rejected or abandoned (e.g., Mikulincer et al., 2003; Wei et al., 2007). In the current study, we chose to use a short version of the ECR (ECR-S) developed by Wei et al. (2007), which is a 12-item scale that captures both Attachment Anxiety (“I want to get close to my partner, but I keep pulling back.”) and Attachment Avoidance (“I try to avoid getting too close to my partner.”). The authors reported adequate reliability (alpha=.78 for Anxiety subscale and .84 for Avoidance subscale) for the ECR-S; the short version of the scale also performed similarly to the original ECR in its test-retest reliability (>.80) and its predictability to relevant criteria such as excessive reassurance seeking, fear of intimacy, depression, and anxiety. Participants rated each item on a scale of 1 ("Agree Strongly") to 7 ("Disagree Strongly"). Their ECR scores were calculated by averaging their ratings across all the items. Higher scores indicate a more secured attachment orientation.

3.3 Participants

I pre-registered the study and aimed to collect data 100-125 participants after exclusion[6]. However, data collection ended prematurely due to COVID-19. A total of 82 participants therefore took part in the current study in exchange of partial course credit for an introductory Psychology course or a compensation of $10. As previously mentioned, participants also earned bonus cash based on their performance in the study tasks. I excluded 5 suspicious participants and 3 inattentive participants (1 for spending less than 250 milliseconds on multiple rating questions for an avatar; 2 for not choosing the required minimum number of 7 self-disclosures to reveal to the avatars). The remaining 74 participants (28 male, 44 female, and 2 non-specified) aged from 18 to 29 (*M*=19.42, *SD*=2.54) and consisted of 29 Asian (39.2%), 25 Caucasian (33.8%), 5 Middle Eastern (6.8%), 5 people of African descent (6.8%), and 10 (13.5%) mixed or other ethnicities.

3.4 Results

3.4.1 Manipulation check

To check whether our manipulation of participants' experiences with the avatars' in the Joint Flanker task effectively differentiated the "good" versus "bad" avatar, I first checked whether participants earned more points when playing with the good than the bad avatar. I fit a linear-mixed model to the total number of points that participants

[6] https://osf.io/axvd5/

earned for Joint Flanker Task, with the partner condition as the predictor, and allowed the intercept to vary across participants. The *lmer* package in R (version 3.6.3) was used. Participants earned significantly more points when playing with the good partner (*M*=136, *SD*=13.7) than with the bad partner (*M*=103, *SD*=15.2), *F*(1, 73)= 402.21, *p*<.001. Moreover, the a linear-mixed model was fit on participants' mood measured by the PANAS after playing with each of the avatars. After controlling for their baseline positive mood measured at the beginning of the experiment, participants reported significantly higher positive affect after playing with the good partner (*M*=60.8, *SD*=21.7) than the bad partner (*M*=56.3, *SD*=22.9), *F*(1, 73)=14.34, *p*<.001. In comparison, their negative affect did not differ as a function of playing with the good (*M*=15.8, *SD*=12.5) versus the bad partner (*M*=16.9, *SD*=12.5), *F*(1,73)=1.50, *p*=.225, after controlling for their baseline negative mood. These findings thus suggested that participants had more positive experiences when playing with the good partner than the bad partner as we intended.

3.4.2 Perceived similarity, perceived knowledge, and liking

To assess participants' perceived similarity and perceived knowledge of the three avatars, a cumulative-link model was fit to the responses on each of these two items using the *ordinal* package in R (version 3.6.3), with partner condition as the predictor, while allowing the intercept to vary across participants. As intended, participants did not perceive significantly different level of similarity among the good (*M*=4.22, *SD*=1.39), bad (*M*=4.10, *SD*=1.35), and neutral avatar (*M*=4.18, *SD*=1.33), χ^2(2)=0.521, *p*=.771. Participants' perceived knowledge of the avatars did vary significantly as a function of our manipulation, χ^2(2)=8.642, *p*=.013. Compared to the neutral partner (*M*=2.50, *SD*=1.21), participants reported knowing the good partner better (*M*=2.92, *SD*=1.41), *b*=0.964, *SE*=0.332, *z*=2.905, *p*=.004, but not the bad partner (*M*=2.68, *SD*=1.16), *b*=0.461, *SE*=0.322, *z*=2.9051.432, *p*=.152.

Next, participants' liking of the avatars was assessed by fitting a linear-mixed model to their aggregated social preferences and trait perceptions of the avatars, which were calculated by averaging the ratings across the corresponding items as in the previous studies. Again, the partner condition was entered as the predictor and the intercept was allowed to vary across participants. Participants showed significantly different social preferences of the avatars, *F*(2, 146)=3.20, *p*=.044. Specifically, participants preferred the good avatar (*M*=3.84, *SD*=1.09) over the neutral avatar (*M*=3.57, *SD*=1.11), *b*=0.270, *SE*=0.117, *t*(146)=2.31, *p*=.022 , whereas preference for the bad avatar (*M*=3.60, *SD*=0.95) did not differ significantly from that of the neutral avatar, *b*=0.030, *SE*=0.117, *t*(146)=0.26, *p*=.795. Participants' trait perceptions of the good (*M*=4.31, *SD*=1.03), bad (*M*=4.28, *SD*=0.94), and neutral avatars (*M*=4.32, *SD*=.95) did not significantly differ from each other, *F*(2, 146)=0.76, *p*=.927. These findings suggest that our manipulation was successful in inducing different levels of liking towards the different avatars, as measured by the social preferences items.

3.4.3 Risk-taking measures: scale reliability and factor analysis

Several different measures of risk-taking were included in this study. Some of these measures evaluate different aspects of risk taking (e.g., behavioural measures versus the self-reported BIS/BAS sensitivity) whereas others assess theoretically similar or overlapped concepts (e.g., the BIS/BAS scales and the SPSRQ). We therefore decided to use the common factors extracted from these measures, instead of all the individual measures, as indicators of participants' risk taking. We first examined the reliability of the self-reported scales and any subscales. A factor analysis was then conducted to extract common factors of the different measures. Lastly, factor scores were stored to represent the distinct aspects of participants' risk taking, which were subsequently used in models that test our hypotheses as described in the next section.

Among the 74 participants included for data analysis, 3 failed the attention check item imbedded in the BIS/BAS scales[7], 1 failed the attention check item imbedded in the

[7] "Please select "Somewhat agree" for this item."

ECR-S scale[8]. These participants' responses on the specific scale for which they failed attention check were coded as missing. No one failed the attention check item in the SPSRQ scale[9].

For the BIS/BAS scales, the BIS subscale (Cronbach's Alpha=0.76) demonstrated adequate internal consistency. In comparison, only one of the three BAS-related subscales, BAS Drive (Cronbach's Alpha=0.77), showed good internal consistency whereas the other two, BAS Reward (Cronbach's Alpha=0.63) and BAS Fun Seeking (Cronbach's Alpha=0.68), showed inadequate consistency. Further examination on the scale items using an Exploratory Factor Analysis with promax rotation revealed that the items on the BIS subscale, the BAS Drive subscale, and the BAS Fun Seeking subscale all loaded on their respective scales. Items on the BAS Reward Responsiveness subscale, however, split between BAS Drive subscale and BAS Fun Seeking subscale[10]. As a result, we grouped these items into the BAS Drive and BAS Fun Seeking items, respectively, instead of treating them as a separate subscale. The revised BAS Drive subscale (Cronbach's Alpha=0.78) and BAS Fun seeking subscale (Cronbach's Alpha=0.72) showed adequate internal consistency. Finally, we calculated participants scores on the BIS subscale, BAS Drive subscale, and BAS Fun Seeking subscale by averaging their ratings cross the corresponding items.

Both the Sensitivity to Punishment (SP) subscale (Cronbach's Alpha=0.83) and the Sensitivity to Reward (SR) subscale (Cronbach's Alpha=0.76) on the SPSRQ achieved adequate-to-good internal consistency; participants ratings on the SP and SR subscales were thus summed, respectively, to obtain their Sensitivity to Punishment and Sensitivity

[8] "Please respond to this item by selecting "4" (Neither agree nor disagree)."

[9] "The University of Western Ontario is located in Canada. Yes=1, No=0"

[10] Items "when I get something I want, I feel excited and energized" and "when I see an opportunity for something I like, I get excited right away" loaded on the BAS Drive factor; items "when I'm doing well at something, I love to keep at it", "when good things happen to me, it affects me strongly", and "it would excite me to win a contest" loaded on the BAS Fun Seeking factor.

to Reward scores. The ECR-S scale also achieved an acceptable level of internal consistency (Cronbach's Alpha=0.70); participants ratings on the ECR-S scale were thus averaged to obtain their ECR scores.

We then proceeded to extract common factors from the different risk measures, including BART, CCT, BIS subscale, BAS- Drive subscale, BAS- Fun seeking subscale, SP subscale, SR subscale, and the ECR-S scale. The Kaiser-Meyer-Olkin factor adequacy (KMO=0.53) suggested that these measures meet the minimal standard for acceptable sampling adequacy (Dziuban & Shirkey, 1974). To extract common factors from the measures, a parallel analysis using principal axis method was carried to determine the number of factors to extract. The parallel analysis suggested three factors, which were consistent with the number of factors with eigen values greater than 1.0. A factor analysis was subsequently conducted to extract three factors. Promax rotation was used to allow the factors to correlate with each other, as all the measures are assessing aspects of the same umbrella term – risk taking. Table 1 presents the factor loadings and correlations among the extracted factors. Together, the three factors explained 49.8% of the variance in the data.

The factor loadings were consistent with the theoretical reasoning behind the measures. The BIS subscale and the SP subscale both loaded positively on the first factor, whereas the ECR-S loaded negatively on this factor. Higher SP and BIS scores reflected individuals' greater motivation to avoid potential punishment, whereas higher ECR-S scores reflect a more secure attachment style and thus lower motivation to avoid interpersonal rejection. The first factor was thus named Punishment Avoidance. The BAS Drive and the SR subscale both assess participants' motivation to seek out and approach potential reward and loaded on the second factor, which was named Reward Approach. The two behavioural measures loaded on the third factor, which was named Behavioral Risk. Interestingly, the BAS Fun Seeking subscale loaded both on the Reward Approach factor and the Behavioural Risk factors, with a slightly heavier loading on the latter. This was reasonable, as the BAS Funk Seeking subscale contains both items that focused on individuals' tendency to seek novelty and potential reward (e.g., "I'm always willing to try something new if I think it will be fun"), which underlies the Reward Approach factor, as well as items that might be assessing participants' impulsivity (e.g., "I often act on the spur of the moment") and sensation seeking (e.g., "I crave excitement and new sensations"), which have been shown to positively related to risk taking tendency as measured in the BART (e.g., Bornovalova et al., 2009; Lauriola, Panno, Levin, & Lejuez, 2014; Lejuez et al., 2002). Factor scores for these three factors, Punishment Avoidance, Reward Approach, and Behavioural Risk, calculated based on Thurstone's regression method (Thurstone, 1934), were stored and subsequently used as indicators of participants' risk taking in the predictive models described in the next section.

Table 2 Experiment 6 Factor Loadings for the Risk Measures

	Factor 1 (Punishment Avoidance)	**Factor 2** (Reward Approach)	**Factor 3** (Behavioural Risk)
Factor Loadings			
BART	0.038	-0.267	**0.682**
CCT	0.173	-0.021	**0.602**
BIS subscale	**0.662**	0.125	0.011
BAS Drive subscale	-0.048	**0.875**	-0.216
BAS Fun Seeking subscale	-0.110	**0.346**	**0.364**
SP subscale	**1.017**	-0.059	0.078
SR subscale	0.080	**0.518**	0.091
ECR-S	**-0.330**	0.080	0.282
Factor Correlations			
Factor 1 (Punishment Avoidance)	-		
Factor 2 (Reward Approach)	-0.136	-	
Factor 3 (Behavioural Risk)	-0.378	0.443	-

Note. Bolded values are factor loadings higher than 0.30. BART= Balloon Analogue Risk Task; CCT= Columbia Card Task; BIS subscale= the Behavioral Inhibition System subscale in the BIS/BAS scales; BAS Drive subscale = the Behavioural Activation System Drive subscale in the BIS/BAS scales (revised)*; BAS Fun Seeking subscale = the Behavioural Activation System Fun Seeking subscale in the BIS/BAS scales* (revised)*;SP subscale= the Sensitivity to Punishment subscale in the SPSRQ; SR subscale= the Sensitivity to Reward subscale in the SPSRQ; ECR-S= the Experiences in Close Relationship Scale- Short form.*

3.4.4 Liking, risk-taking, and self-disclosures

To properly test my hypotheses, two-level linear-mixed models should be used to examine whether (1) at the within-subject level, greater liking led to greater self-disclosures to the avatars (level-1 effect); (2) at the between-subject level, greater risk-taking tendency led to greater overall self-disclosure (level-2 effect); (3) there was an interaction between BIS/BAS and liking on self-disclosure to the avatars (cross-level interaction effect). However, because the data collection was ended prematurely, our actual sample size was considerably smaller than planned. Insufficient level-2 clusters (i.e., number of participants in this study) can lead to an elevated nonconvergence rate for linear-mixed models (Maas & Hox, 2005) as well as a decreased statistical power to detect effects, especially for cross-level interactions (Arend & Schäfer, 2019). Therefore, to maximize power, I used random-intercept-only models instead of allowing level-1 effects to vary across individual participants (i.e., random slopes) (Matuschek et al., 2017) and subsequently refrained from testing any cross-level interaction effects between BIS/BAS sensitivity and liking. We will resume the data collection process once the situation allows and test for interaction effects when adequate data become available.

Following the step-wise approach to model building proposed in Hox (2010) (p56-59), an intercept-only model was first analyzed, allowing the examination of the Intraclass Correlation (ICC) of the self-disclosure data. Next, the level-1 predictors were included in the model to examine the contribution of the level-1 factors, or avatar-level differences, in predicting participants' self-disclosures. For the level-1 predictors, I included participants' social preferences of the avatars as the indicator for their liking of the avatars. I also included their perceived knowledge and perceived similarity as two additional level-1 predictors to examine whether liking predicts self-disclosures after controlling for these two factors. Finally, the level-2 predictors were included in the model to examine how participant-level individual differences predict participants' self-disclosures. In addition to the three risk-taking factors extracted from the risk-taking measures, as described in the last section, I also included participants' gender and age as level-2 predictors to control for the effects of these two factors. All the predictor variables were grand-mean centered to assist the interpretation of the results, with the exception of the risk factor scores which were already standardized scores. Each later model was compared with the previous to examine whether including the additional predictors improved the model fit. To enable model comparison, I only included participants with complete data on all the variables included in the models (N=68): the four participants who failed any of the attention check imbedded in the questionnaire measures were excluded; two additional participants were excluded for missing gender information. This model building process was repeated for the three self-disclosure measures included in the current study, namely, participant's number of self-disclosures to each avatar, their proportion of deep self-disclosures to each avatar, and the number of extra self-disclosures that they made to each avatar.

I first examined the effects of the predictors on the number of self-disclosures that participants made to the avatars. See Table 2 for the coefficients and their significance level across three models. The intercept in Model 0 suggested that participants on average self-disclosed 9 statements to the avatars. Furthermore, the variance in subject-level intercept was very large in comparison to the residual error, resulting in an ICC of 0.81[11], which suggested that 81% of the variance in the participants' number of self-disclosures to the avatars could be traced to subject-level differences. It was thus not surprising that none of the within-subject level-1 factor significantly predicted the number of self-disclosures. In addition, participants' number of self-disclosures did not seem to be predicted by the risk-taking factors or their age or gender. Comparing the model deviance revealed that adding in the level-1 predictors did not improve the fit beyond the intercept-only model, $\chi^2(3)=2.54$, $p=.468$; nor did adding in the level-2 predictors improve the fit beyond that with level-1 predictors only, $\chi^2(5)=6.95$, $p=.224$.

[11] ICC= Subject-level variance / (Subject-level variance + residual error variance) (Hox, 2010)

Table 3 Experiment 6 Model Comparisons for Number of Self-Disclosures

		M0: intercept only	M1: level-1 predictors only	M2: all predictors
Fixed Part	Intercept	9.004(0.33)***	9.011(0.33)***	9.065(0.34)***
	Social Preferences		-0.179(0.20)	-0.248(0.20)
	Perceived Knowledge		-0.067(0.13)	0.095(0.13)
	Perceived Similarity		-0.060(0.11)	-0.042(0.12)
	Punishment Avoidance			0.146(0.34)
	Reward Approach			0.574(0.39)
	Behavioural Risk			-0.538(0.42)
	Gender			-0.358(0.68)
	Age			0.071(0.15)
Random Part	Subject-level variance	6.779(2.60)	7.033(2.65)	6.916(2.63)
	Residual errors	1.598(1.26)	1.577(1.26)	1.571(1.25)
Deviance		851.66	849.13	842.18

*Note. *: significant at p=.05 level; **: significant at p=.01 level; ***: significant at p=.001 level*

Next, we examined effects of the predictors on the proportion of deep self-disclosures that participants made to each avatar. As shown in Table 3, the intercept-only model suggested that, on average, 45% of the statements that participants chose to self-disclose to the avatars were on the high-depth topics. The subject-level variance in the intercept was still relatively large compared to the variance of the residual, resulting in an ICC of 0.66, which suggested that 66% of the variance in participants' proportions of deep self-disclosures was at the subject level. Including the level-1 predictors significantly improved the model fit, $\chi^2(3)=10.40$, *p*=.015. Model 1 results suggested that, as we hypothesized, greater preference for an avatar predicted a higher proportion of deep self-disclosures, *b*=0.033, *SE*=0.014, *t*(194.83)=2.807, *p*=.006. Specifically, when holding other variables else constant, an increase of one point on participants' social preference rating for an avatar was associated about 3% increase in the proportion of deep self-disclosures to that avatar. Interestingly and surprisingly, perceived similarity also significantly predicted the proportion of deep self-disclosures to the avatar, but in the opposite direction, *b*=-0.020, *SE*=0.007, *t*(177.80)= -2.904, *p*=.004. Including the level-2 predictors further improved the model fit, $\chi^2(5)=11.08$, *p*=.050. Here, the only risk-taking factor that significantly predicted participants' proportions of deep self-disclosures was their scores on the Behavioural Risk factor. Again, surprisingly and opposite to our prediction, higher scores on Behavioural Risk propensity *negatively* predicted the proportion of the deep self-disclosures, *b*=-0.045, *SE*=0.018, *t*(61.93)=-2.589, *p*=.012. As the factor scores were standardized, the coefficient for Behavioural Risk propensity suggested that with each 1 standard deviation increase in Behavioural Risk propensity, participants' mean proportion of deep disclosures to the avatars would *decrease* by 4.5%. In other words, the higher participants were in risk taking as represented by their scores on the Behavioural Risk factor, the more superficially they self-disclosed to the avatars.

Table 4 Experiment 6 Model Comparisons for Proportion of Deep Self-Disclosures

		M0: intercept only	M1: level-1 predictors only	M2: all predictors
Fixed Part	Intercept	0.454(0.01)***	0.453(0.01)***	0.456(0.01)***
	Social Preferences		0.033(0.01)**	0.033(0.01)**
	Perceived Knowledge		0.0004(0.01)	0.001(0.01)
	Perceived Similarity		-0.019(0.01)**	-0.020(0.01)**
	Punishment Avoidance			0.017(0.01)
	Reward Approach			-0.030(0.02)
	Behavioural Risk			-0.045(0.02)*
	Gender			-0.035(0.03)
	Age			-0.005(0.01)
Random Part	Subject-level variance	0.013(0.11)	0.012(0.11)	0.011(0.10)
	Residual errors	0.006(0.08)	0.006(0.08)	0.006(0.08)
Deviance		-321.39	-331.79	-342.87

*Note. *: significant at p=.05 level; **: significant at p=.01 level; ***: significant at p=.001 level*

Finally, we examined the effects of the predictors on the number of extra self-disclosures that the participants chose to type in the open-ended textbox for any of the avatars. The ICC of the responses was 0.77, suggesting that 77% of the variance in the number of extra self-disclosures resulted from between-subject differences. As can be seen in Table 4, none of the level-1 or level-2 predictors significantly influenced this outcome variable. Including the level-1 factors in the model did not improve model fit beyond the intercept-only model, $\chi^2(3)=5.72$, $p=.126$, nor did including the level-2 factors improve the fit over the model with only level-1 predictors, $\chi^2(5)=3.64$, $p=.603$.

Table 5 Experiment 6 Model Comparison for Number of Extra Self-Disclosures

		M0: intercept only	M1: level-1 predictors only	M2: all predictors
Fixed Part	Intercept	0.299(0.05)***	0.296(0.05)***	0.315(0.05)***
	Social Preferences		0.047(0.03)	0.038(0.04)
	Perceived Knowledge		0.031(0.02)	0.036(0.02)
	Perceived Similarity		-0.035(0.20)	-0.034(0.02)
	Punishment Avoidance			0.006(0.05)
	Reward Approach			0.023(0.060)
	Behavioural Risk			-0.023(0.06)
	Gender			-0.173(0.10)
	Age			0.002(0.02)
Random Part	Subject-level variance (in intercept)	0.163(0.40)	0.155(0.39)	0.160(0.40)
	Residual errors	0.049(0.22)	0.049(0.22)	0.049(0.22)
Deviance		125.02	120.02	116.38

*Note. *: significant at p=.05 level; **: significant at p=.01 level; ***: significant at p=.001 level*

3.5 Discussion

In Experiment 7, we examined how participants' self-disclosures to different social partners were influenced by their liking of these partners and their own risk-taking tendencies. We found some support for Effect 2 in Collins and Miller (1994). Specifically, participants' social preferences for the avatars significantly and positively predicted the proportion of deep self-disclosures that they chose to share with the avatars. Increased liking of the avatars, however, did not predict the number of self-disclosures they chose to share or the number of extra self-disclosures that they made to the avatars. Thus, contrary to prediction, participants' risk-taking tendencies, as represented by their scores on the Behavioural Risk propensity, *negatively* predicted their proportions of deep self-disclosures to the avatars. Again, this effect was not observed on the total number of self-disclosures or the number of extra self-disclosures. Another surprising finding was that perceived similarity also negatively predicted the proportion of deep self-disclosures. Participants perceived knowledge of the avatars, their age and gender, and their scores on the Reward Sensitivity and Punishment Sensitivity factors did not significantly influence participants' self-disclosures.

Before discussing these results, I wish to remind readers that the present sample was considerably smaller than planned and this might have resulted in various issues regarding the results. First, because of the sample size limitation, there might have not been adequate power to detect some effects, especially second-level effects. On the other hand, the small sample size, as well as the random-intercpt-only models used in analyzing the data, might have led to Type I errors such that the significant effects that we did find might not reflect effects that truly exist (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). Furthermore, because we refrained from testing any cross-level interactions, we do not have information regarding how these effects interact with each other, which might alter our interpretations of the results.

With these precautions in mind, are there possible mechanisms that might explain why participants self-disclosed more superficially when they had higher risk-taking tendencies, or perceived an avatar to be more similar to themselves? One possibility is that those who scored high on the Behavioral Risk propensity might perceive a lower general risk of being rejected by another person. As risk perceptions are frequently found to be negatively associated with risk-taking decisions (Mills et al., 2008; Ryb et al., 2006), higher scores on the Behavioural Risk propensity might reflect a general low perception of risk. Thus, these participants might not experience as strong of a need to signal their liking and interest in affiliation by using deep self-disclosures, compared to those who are low on risk-taking and perceive a greater risk of rejection. That is, these participants may worry less about a particular individual declining to reciprocate a friendship. Similarly, participants who perceived the partner as more similar to themselves might expect a greater friendship intention from the partner, as we found in the previous experiments. As a result, they might also expect the social partner to accept them, reducing their need to use deep self-disclosures to convey their liking to the other.

Another possibility has to do with the specific self-disclosure statements used in the experiment. Consistently with social penetration theories, the high self-disclosure items used in our experiments consisted of topics such as values and core beliefs (e.g., best quality in a friend; most valued thing in life), which might have been viewed as abstract and serious by the participants. In comparison, the low self-disclosure items were consisted of topics that might be considered as less abstract, and thus, potentially more fun (e.g., places to travel to; hobby to pick up; favorite food). Considering that participants' scores on the BAS- Fun Seeking subscale also loaded on the Behavioural Risk factor, it was possible that a high score on this factor also reflected participants' sensation seeking and desire for excitement. Thus, these participants might have self-disclosed more superficially simply because these statements were considered as more fun than the statements consisting of deeper self-disclosures.

Without further evidence, both these possibilities discussed here are speculations. Additional data and replications are needed to provide cummulative evidence for these ideas, as in Experiments 1 through 5, to enhance our confidence in our conclusions regarding the role played by one’s liking of a social partner and their risk-taking tendencies on their self-disclosures to the partner. Further studies are thus needed to determine whether these surprising patterns can be replicated and what the mechanisms underlying these surprising patterns might be.

Chapter 4

4 Experiment 7: Negative Perceptions of a Self-Disclosing AI: the Potential Role of the Uncanny Valley Effect

Recent years have witnessed increasing involvement of artificially intelligent (AI) agents and systems in our daily lives. Many such AIs are designed to socially interact with people in different settings, such as health care and companionship for the elderly (Robinson et al., 2014), learning and education (Mubin et al., 2013), and mental health screening (Lucas et al., 2014). Identifying factors that may help or hinder people's perceptions of these social AIs is crucial. Such knowledge will help people design AIs in a way that facilitates positive human-AI relationships and, ultimately, the best utilization of the services these AIs offer.

Despite the key roles of self-disclosure in typical social relationship development, few studies have investigated how people perceive a self-disclosing AI, especially one that discloses some level of similarity to themselves. On the one hand, some researchers have shown that people follow human social norms when interacting with computers and AIs (Moon, 2000; Moon & Nass, 1996; Nass & Moon, 2000). People might thus form positive perceptions of an AI that self-discloses similarity to themselves just as they do with human partners. One the other hand, an automatic agent that is high in human realism might induce eerie and unnerving feelings, known as the "uncanny valley" effect (Mori, 1970). Given that self-disclosing behaviour might be seen as highly human-like, an AI who self-discloses, especially one that reveals a high level of similarity to oneself through its self-disclosures, might fall into the uncanny valley and elicit negative perceptions from the human user. Here, we explored people's perceptions of a self-disclosing AI social partner, in comparison to a human partner, with varying degrees of similarity to oneself.

4.1 Rationale

To date, the question of how human users respond to an AI's self-disclosures has received little attention. There is, however, evidence suggesting that an AI's self-disclosures can be used to facilitate users' social reactions towards it. For example, people were found to like an AI agent more when it self-disclosed at a deeper rather than superficial level about itself (Kang & Gratch, 2011). People also disclosed more personal information to a computer that disclosed to them first than to one that did not (Moon, 2000). An AI that self-disclosed human "back stories" as its own history was preferred over one that told the same stories as if describing another human (Bickmore et al., 2009). As we found in the previous studies, people like others who share and disclose a high level of similarity to themselves. Could self-disclosed similarity from an AI partner enhance the human user's positive perceptions and liking of the AI?

4.1.1 Similarity-attraction or the uncanny valley effect?

As described in the previous chapters, people tend to like similar others more than they do dissimilar others (e.g., Byrne, 1971; Montoya et al., 2008) and perceive more similar others as socially warmer and more intellectual (Lydon et al., 1988). Some evidence suggests that the similarity-attraction effect also applies to human-AI interactions (de Melo et al., 2014; Moon & Nass, 1996; Verberne et al., 2015). For instance, participants rated a computer that displayed a similar submissive-dominant personality trait to themselves as friendlier, smarter, and more helpful than a computer that displayed the opposite trait (Moon & Nass, 1996). People also showed greater trust of a virtual driver who was more similar to themselves in a driving simulation study (Verberne et al., 2015). The researchers made the virtual driver look more or less like the participant by morphing a default digital face with either the participant's own photo or another person's photo. The virtual driver also mimicked the participant's head movements to different degrees. The participants reported greater trust to the virtual driver that looked and behaved more like them. The similarity-attraction effect also emerges when people perceive the AI to be in the same social category as themselves: participants showed greater trust in and allocated more resources to computer agents that appeared as from the same ethnicity group as themselves (de Melo et al., 2014). Furthermore, people showed in-group bias in favor of a computer agent when being paired with the agent in an arbitrary group (de Melo et al., 2014). It is therefore possible that increased self-disclosed similarity from an AI partner would lead to more positive social outcome in human-AI interactions as it does in human-human interactions.

However, an AI agent's greater similarity to human may not always lead people to perceive it more positively. The "uncanny valley" effect (Mori, 1970) describes a phenomenon in which people's sense of familiarity towards a robot initially increases as people perceive greater similarity to the robot but sharply drops when the agent or robot becomes too human-like. Mori (1970) argues that people would experience a sudden loss of familiarity when they notice the subtle differences between a human-like agent and a real human, which induce unnerving and eerie feelings towards the agent. The uncanny valley effect has been observed in a considerable number of empirical studies (e.g., Ferrari, Paladino, & Jetten, 2016; Kätsyri, Förger, Mäkäräinen, & Takala, 2015; Mitchell et al., 2011; Tinwell, Grimshaw, Nabi, & Williams, 2011). The exact nature of the uncanny valley effect (Bartneck et al., 2007) and the conditions necessary to create it (e.g., MacDorman, 2006) are still cause for debate. Some researchers suggested that the uncanny valley might be a result of inconsistency of human realism such as different levels of human-likeness in facial features (K. F. MacDorman & Chattopadhyay, 2016; Seyama & Nagayama, 2007) and audio-visual asynchrony (Tinwell et al., 2015). Others have provided a higher-level explanation for the uncanny valley and suggested that it raises from a blurred distinction between human and machines and a perceived threat to human uniqueness (Ferrari et al., 2016).

Interestingly, even though most empirical studies on the uncanny valley effect have focused on agents' physical appearance (e.g., MacDorman & Chattopadhyay, 2016; Seyama & Nagayama, 2007) or movement (e.g., Saygin, Chaminade, Ishiguro, Driver, & Frith, 2012), some recent work suggests that if an agent appears too human-like in its mental capacities, it may also induce such uncanny and eerie feelings (K. Gray & Wegner, 2012; Stein & Ohler, 2017). Following the reasoning of the uncanny valley effect, self-disclosed similarity from an AI might potentially lead to negative reactions of the human user to the AI.

4.1.2 Perceptions, dehumanization, and anthropomorphism

One key variable in the uncanny valley effect is how human-like an agent appears to be, which is frequently termed "anthropomorphism" in the human-AI interaction literature. Anthropomorphism refers to people's tendency to attribute human characteristics, intentions, and motivations to non-human objects (Epley et al., 2007). It is important to differentiate between *anthropomorphic features* and *anthropomorphism*: the former refers to the agents' objective human-like features, such as facial appearance and movements, whereas the later describes the subjective experience of the user and how they personally view the agent as human-like. While anthropomorphic features may lead to anthropomorphism (e.g., Hegel, Krach, Kircher, Wrede, & Sagerer, 2008; Mitchell et al., 2011), the extent to which people "humanize" an agent is also heavily influenced by social-cognitive processes. For example, people are more likely to anthropomorphize a non-human object or animal when they are socially motivated (Epley et al., 2007).

Dehumanization occurs when people deny a target human attributes, which may be viewed as the converse of anthropomorphism (Haslam, 2006). Specifically, Haslam (2006) articulated two types of dehumanization, animalistic dehumanization and mechanistic dehumanization. Haslam suggested that if others are denied "Human Uniqueness" such as civility, refinement, and moral sensibility, they are viewed like animals, whereas when they are denied "Human Nature" such as emotional responsiveness, interpersonal warmth, and individuality and agency, they are often viewed like machines. We therefore argue that, because of the unpleasant and eerie feelings experienced in the uncanny valley effect, people might view the agent as lacking interpersonal warmth, leading to decreased social motivation to interact with it. As a result, the agent might be dehumanized, or less anthropomorphized, and viewed as more machine-like, even though its anthropomorphic features and human-likeness trigger the uncanny and unpleasant feelings in the first place.

4.2 Overview of the current experiment

In this experiment, we were interested in exploring people's perceptions and liking of a self-disclosing AI social partner, compared to a human partner, with different levels of similarity to themselves. We manipulated the AI's self-disclosed similarity through a conversational virtual human and measured participants' perceptions of it. Moreover, to compare participants' perceptions of an AI versus another human partner, half of the participants were made to believe that the virtual human was an AI while the rest believed that it was controlled by another human. In this way, we independently manipulated the partner's self-disclosed similarity to the participants, as well as the perceived identity of the partner.

We measured participants' social perceptions of their partner, including how warm and competent they perceived their partner to be (Fiske et al., 2007), as well as how much they attributed the capability to experience feelings and exhibit agency (H. M. Gray et al., 2014), both considered characteristics that separate humans from machines (Haslam, 2006). Participants also reported how much they see their partner as a machine versus a human, or their level of anthropomorphism of their partner (Bartneck et al., 2009). In addition, participants self-reported their liking of and perceived rapport with their partner. Finally, we included two behavioral measures, the length of the conversation between the participant and the virtual human and participants' facial expressions during the conversation, to provide some validation to our self-report measures.

4.3 Hypotheses

We anticipated a main effect of partner identity, such that the human partner would be perceived as warmer, more competent, more capable to experience feelings and exhibit agency, and more human-like than the AI partner. We also expected people to like the human partner more than the AI partner.

As to the effect of self-disclosed similarity, the similarity-attraction effect and the uncanny valley effect would generate different predictions. Specifically, the similarity-attraction effect would suggest a main effect of similarity, such that greater self-disclosed similarity would be associated with more positive perceptions and greater likings for both the human and the AI partner. In contrast, the uncanny valley effect would suggest an interaction between partner identity and self-disclosed similarity. Because increased similarity would lead to more positive outcome in human-human interaction but more negative outcome in human-AI interaction, participants would favor the human partner over the AI partner to the greatest extent in the high similarity, followed by the medium similarity, and least in the low similarity condition.

4.4 Method and materials

4.4.1 Participants

A total of 195 participants (74 Male, 119 Female, 2 declined to state) were recruited from the Los Angeles area. Sixty of the participants were recruited online via Craigslist and the remaining participants were recruited in a university in Los Angeles. Informed consent was obtained from all participants. Study sessions were conducted in a laboratory setting.

4.4.2 Procedure

Each participant had a conversation with a virtual human named Julie (Artstein et al., 2016) displayed on a 30-inch computer monitor in an individual lab room. To manipulate the perceived identity of their conversation partner, half the participants (randomly selected) were told that Julie was remotely controlled by another person in a different room, while the rest were told that Julie was an AI agent controlling its own behaviours (i.e., Partner Identity: Human vs. AI). In fact, Julie was always remotely controlled by me through a “Wizard of Oz” style system (Artstein et al., 2016). I observed participants from a different room via real-time video streaming and interacted with them using the “Wizard of Oz” control panel (Figure 4-1). Specifically, I was able to control Julie’s verbal responses by clicking pre-selected options on the control panel. Figure 4-1 shows a screenshot of Julie and the control panel that generated Julie’s verbal responses.

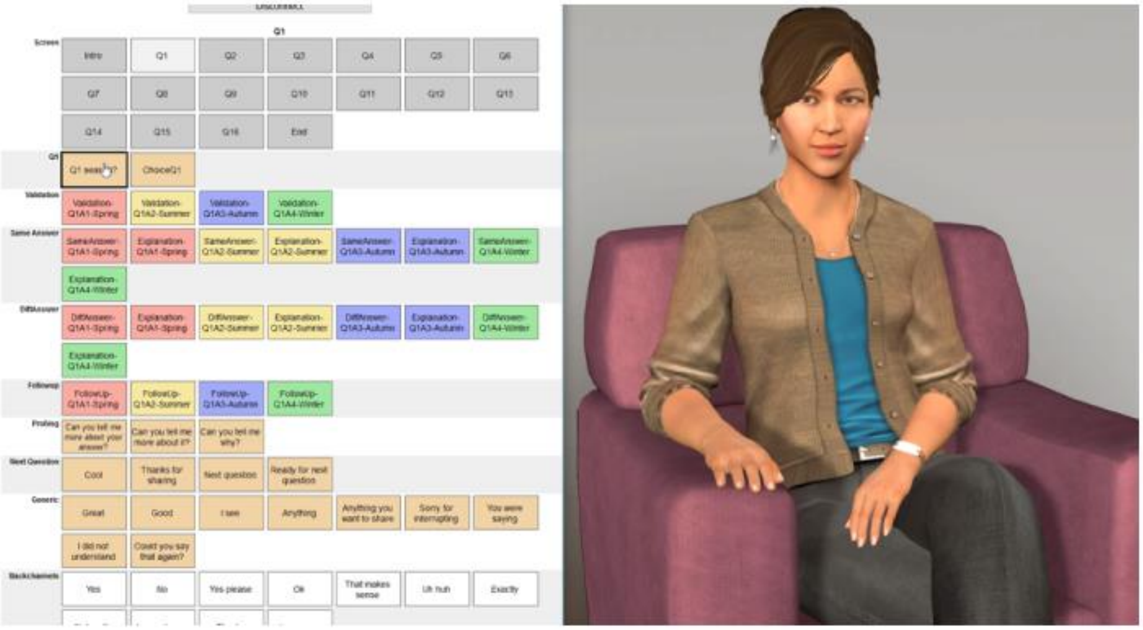


Figure 4-1 A screen shot of Julie (right; Artstein et al., 2016) and the “Wizard of Oz” control panel used in Experiment 7 (left).

During the conversation, Julie asked the participant 16 multiple-choice questions about their attitudes and preferences on various topics. These questions were selected from the 40 multiple-choice questions that we used in the previous experiments. Half these questions were relatively superficial (e.g., “What is your favorite cuisine: Italian, Japanese, Chinese, or Mexican?”) while the remaining were relatively personal (e.g., “Which of these things is most stressful to you: working, thinking about future, social conflict, or unplanned events?"). To make the conversation seem naturalistic, Julie asked the questions in a gradually more personal manner (Moon, 2000). The question order was thus fixed for all participants.

Upon receiving the participant's response to a question, Julie provided a positive validation to their answer (e.g., "Italian dishes are delicious!") before giving the participant its own answer and a justification for the answer (e.g., "Japanese is my favorite. Japanese food always looks pretty and tastes fresh! "). Julie then asked the participant to elaborate on their answer to that question (e.g., "What are the things you like about Italian food?") and moved on the next question. The entire conversation usually lasted 10 to 15 minutes.

To manipulate the similarity between Julie and the participants, we varied the number of questions in which Julie's answer was the same as a participant's own choice. For participants in the high-similarity condition, Julie's answer was the same as their own for 12 out of the 16 questions. Julie's answer was the same as the participant's own for 8 of 16 questions in the medium-similarity condition and 4 out of 16 questions in the low-similarity condition. We randomized the specific questions for which Julie provided the same versus different answers across participants, as well as the specific answer that Julie provided when its answer differed from the participant's[12].

After the conversation, participants completed the manipulation checks and the dependent measures in the same room on a laptop. The entire session was video recorded to obtain participants' facial expression data during the conversation with Julie.

[12] A master sheet that contains the randomized trials and answers was generated using this python code: https://github.com/yixian625/research/blob/master/trial%20numbers%20and%20answers%20generator_AI%20self-disclosure%20project.py. The experimenter followed the master sheet to determine how to respond to each participant in each trial.

4.4.3 Measures

4.4.3.1 Manipulation checks

Two items were used to check the effectiveness of the manipulations. Participants were asked to indicate whether they thought Julie's behaviours were controlled by a human or an AI. Those who failed this check were subsequently removed from all the data analyses. Participants were also asked to rate how much they thought they and their partner were alike on a scale of 1("Very much different") to 7("Very much alike").

4.4.3.2 Warmth and competence scale

Participants' perceptions of their partner's warmth and competence were measured using items modified from Fiske, Cuddy, Glick, and Xu (2002). They rated the following attributes in reference to their conversation partner on a 7-point scale (1 "Not at all" to 7 "Very much"): sociable, likable, good-natured, tolerant, friendly, sincere, trustworthy, well-intentioned, pleasant, and warm (Warmth items); organized, confident, capable, efficient, independent, competent, expert, competitive, skillful, and intelligent (Competence items).

4.4.3.3 Mind perception scale

Participants reported mind perceptions of their partners on two dimensions, the experience dimension (MP-E) and the agency dimension (MP-A), using the Mind Perception Scale (H. M. Gray et al., 2014). The MP-E dimension reflects how much the participants perceived their conversation partner to be able to experience feelings and emotions, whereas the MP-A dimension reflects how much they perceived the partner to be able to think and act (H. M. Gray et al., 2014). Participants rated on a 7-point scale (1 "Not at all" to 7 "Very much") how much their partner appeared to be capable of experiencing the following: hunger, joy, fear, pain, pleasure, rage, desire, personality, consciousness, pride, and embarrassment (the MP-E items); self-control, morality, memory, emotion recognition, planning, communication, and thought (the MP-A items).

4.4.3.4 Anthropomorphism scale

To assess how much participants perceived their conversation partner as human-like, participants filled out the Anthropomorphism subscale of the Godspeed Questionnaires (Bartneck et al., 2009). Participants rated their impressions of the conversation partner on the following pairs of attributes: Artificial versus Lifelike, Fake versus Natural, Unconscious versus Conscious, Moving Rigidly versus Moving Elegantly, Machine Like versus Human Like, with each attribute of a pair appeared on one end of a 7-point scale.

4.4.3.5 Liking and rapport scale

Participants reported their liking of and rapport with their conversation partner on 9 questions adapted from Gratch, Wang, Gerten, Fast, and Duffy (2009). They rated their liking and trust towards their conversation partner as well as their enjoyment of the interaction on a 7-point scale (1 “Not at all” to 7 “Very much”). Sample items include “How much did you like your interaction partner?”, “How much did you enjoy the interaction?”, and “How much rapport did you feel with your interaction partner?”.

4.4.3.6 Conversation length

We extracted the length of the conversation from the session video recorded for each participant. It was assumed that participants would engage in longer conversations with Julie if they were enjoying the interaction more.

4.4.3.7 Facial action unit activities

We analyzed participants’ facial expressions captured in the session videos (29.97 frames/second) during their conversation with Julie. We used iMotion’s Emotient FACET module (iMotions, 2016) and focused on three Action Units (AU) defined in the Facial Action Coding System (FACS; Ekman & Friesen, 1971; Ekman & Rosenberg, 1997): evidence for AU4 Brow Lowerer/Frowning, AU6 Cheek Raiser, and AU12 Lip Corner Puller were analyzed[13]. AU6 and AU12 are commonly considered as indicators of smiling and have been found to correlate with self-reported happiness (Ekman et al., 1980). AU4 has frequently been used in coding negative emotions such as anger, fear, and sadness (e.g., Ekman & Rosenberg, 1997; iMotions, 2016). We expected that participants who enjoyed interacting with Julie more would show greater activity in AU6 and AU12 and less activity in AU4 during the conversation.

4.5 Results

We excluded 40 participants from the subsequent data analyses for failing the partner identity manipulation check. The remaining 155 participants (96 female, 57 male) consisted of 67 (43%) White, 16 (10%) people of African descent, 36 (23%) Asian, and 36 (23%) Other or Mixed ethnicity individuals. Most participants (74%) fell into the 18-24 age group and the oldest participant fell into the 65-74 age group. One participant did not consent to any use of their video data so those data were subsequently excluded from analyses of conversation length and facial expressions.

All the dependent variables were subject to a 3 (Similarity: High vs. Medium vs. Low) X 2 (Partner Identity: Human X AI) ANOVA. See the **Appendix C** for the detailed ANOVA results on all the measures.

[13] The AU activity values generated by FACET are similar to z-scores such that the positive values indicate presence of an AU activity and negative values indicated absence of an activity. As it is not readily interpretable how different negative values reflect different degrees of absence, we recoded all negative values into 0 before analyzing the AU activity.

4.5.1 Experimenter bias check

To check for potential experimenter bias, the number of responses that Julie gave to each participant was extracted from the session log files. An ANOVA on the number of responses did not yield any significant partner or similarity main effects or interactions, suggesting that Julie's responses to participants were relatively consistent across conditions.

4.5.2 Similarity manipulation check

An ANOVA on the similarity manipulation check item revealed a significant main effect of similarity, $F(2, 149)=8.61$, $p<.001$, $\eta^2=.104$. Participants in the High Similarity condition ($M=5.02$, $SD=1.38$) rated their conversation partner as more similar to themselves than did those in the Medium Similarity condition ($M=4.22$, $SD=1.40$) and the Low Similarity condition ($M=3.93$, $SD=1.39$). No other significant effects were found on this item. This finding suggests that our similarity manipulation was successful.

4.5.3 Warmth and competence scale

We first conducted a factor analysis on all the items to examine whether they load on their respective Warmth and Competence subscales. Principle Component Analysis and Promax Rotation were used, allowing the extracted factors to correlate with each other (Osborne et al., 2008). Two positively correlated ($r=.643$) factors were extracted from the data, with all the Warmth items loading on one and all the Competence items loading on the other. We thus calculated the aggregated means for the Warmth items and the Competence items respectively. An ANOVA test on the mean Warmth scores did not find significant main effect of Similarity or Partner Identity. However, a significant Similarity X Partner Identity interaction effect was found, $F(2, 149)=4.38$, $p=.014$, $\eta^2=.056$. Specifically, pair-wise comparisons suggested that the human partner was rated ($M=5.54$, $SD=0.85$) as significantly warmer than the AI partner ($M=4.64$, $SD=1.46$) in the *Medium Similarity* condition; the difference was not significant in the other two

Similarity conditions (Figure 4-1a). See the Appendix C for the mean and SD for each cell and the significant levels for the pair-wise comparisons for all dependent measures

An ANOVA on the Competence scores found no significant main effects for Similarity or Partner Identity. Even though the overall interaction effect was not statistically significant, *F*(2, 149)=2.32, *p*=.102, η^2=.030, pair-wise comparisons showed the same pattern where the human partner (*M*=4.96, *SD*=0.90) was perceived as significantly more competent than the AI partner (*M*=4.20, *SD*=1.57) (Figure 4-1b).

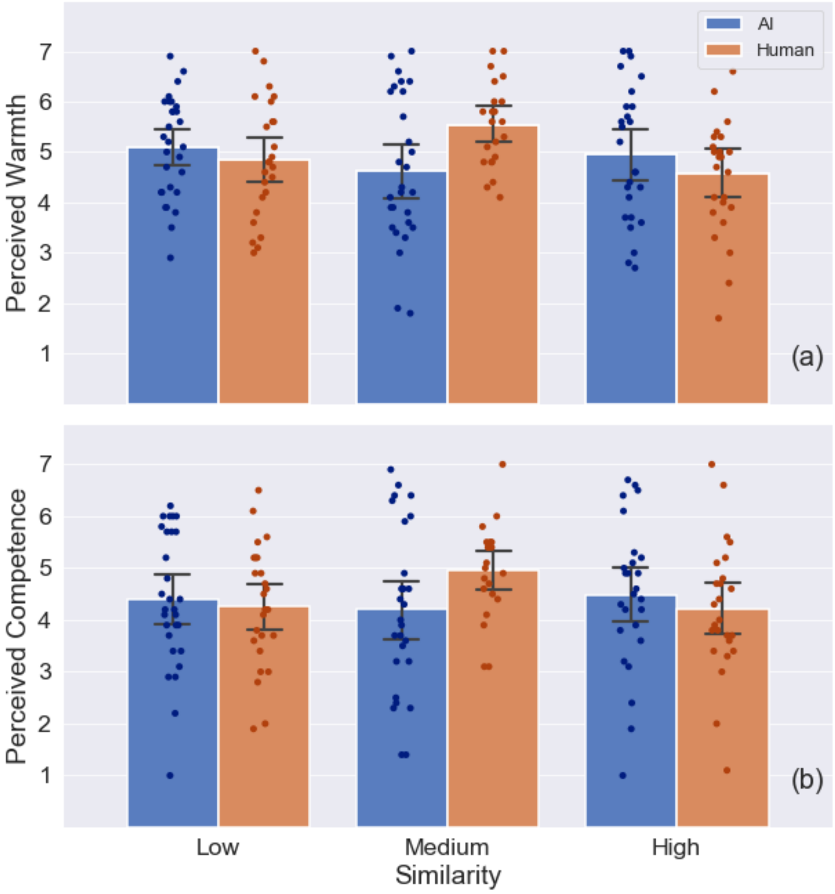


Figure 4-2 Bar plots for participants' ratings on the partner's perceived warmth (a) and perceived competence (b) in Experiment 7. The dots represent the individual data points. Error bars represent 95% CI.

4.5.4 Mind perception scale

We conducted a factor analysis on the Mind Perception Scale items to examine whether the MP-A and MP-E items loaded on their corresponding factors. Principle Component Analysis and Promax Rotation extracted two factors. Most items loaded on their corresponding factors with two exceptions: although Personality and Consciousness were theorized as MP-E attributes (H. M. Gray et al., 2014), they loaded instead with the other MP-A items (Personality loading=.56; Consciousness loading=.71) in our data. As

Personality and Consciousness could also be viewed as reflecting one's abilities to act and think, we subsequently considered them as MP-A items. The two extracted factors correlated highly with each other (*r*=.640), suggesting that the conversation partner's perceived ability to experience feelings and emotions go hand in hand with their perceived ability to think and act.

The aggregated mean scores for MP-A and MP-E were calculated respectively to examine any effects of our manipulations. An ANOVA on the mean MP-A scores revealed a significant Similarity X Partner Identity interaction, *F*(2, 149)= 4.14, *p*=.018, η^2=.053 (Figure 4-2a), and no significant main effect of Similarity or Partner Identity. Pairwise comparisons found that the human partner was perceived (*M*=4.78, *SD*=0.90) to be significantly more capable of thinking and acting than the AI partner (*M*=3.64, *SD*=1.43) in the Medium Similarity condition, but not in the other two Similarity conditions.

An ANOVA on mean MP-E scores also revealed a significant Similarity X Partner Identity interaction effect, *F*(2,149)=3.67, *p*=.028, η^2=.047, with the same pattern where the human partner (*M*=3.14, *SD*=1.05) was perceived as more capable of experiencing feelings and emotions than the AI partner (*M*=2.21, *SD*=1.26), but only in the Medium Similarity condition (Figure 4-2b). No significant main effects were found.

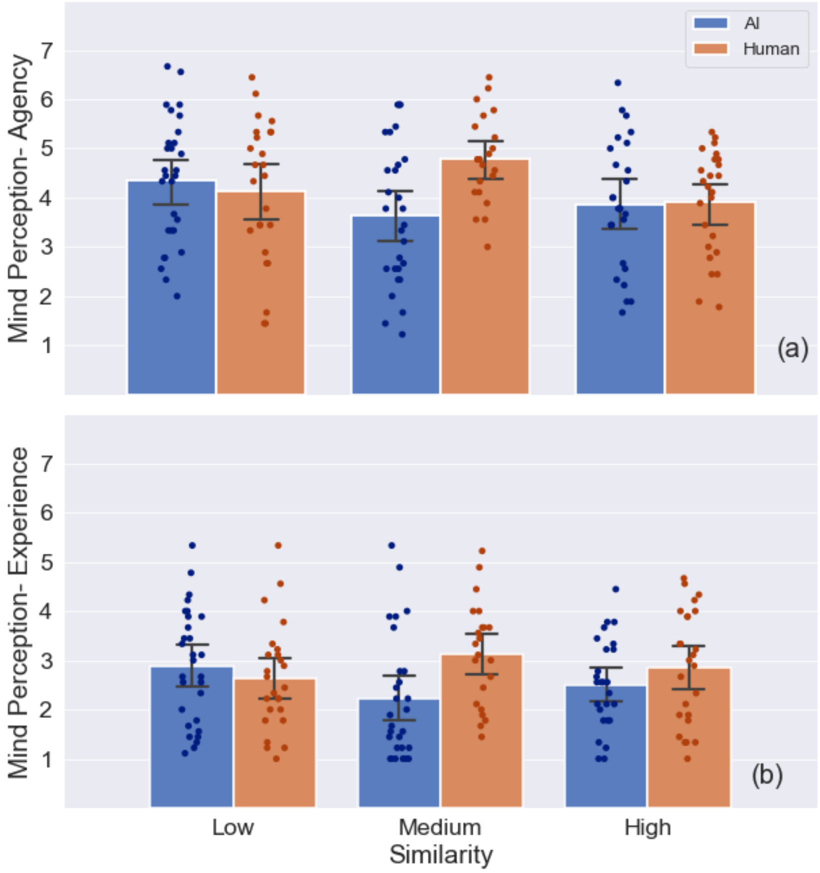


Figure 4-3. Box plots for participants' ratings on the partner's Mind Perception- Agency dimension (a) and Mind Perception- Experience dimension (b) in Experiment 7. The dots represent the individual data points. Error bars represent 95%.

4.5.5 Anthropomorphism scale

The Anthropomorphism Scale items showed a high scale reliability, *Cronbach's Alpha* =.880. An ANOVA on the aggregated mean scores of Anthropomorphism found a borderline significant effect for the Similarity X Partner Identity interaction, *F*(2, 149)= 2.83, *p*=.062, η^2=.037. As above, pair-wise comparisons revealed that indicated a marginally higher level of anthropomorphism towards their human partner (*M*=3.48, *SD*=1.31) than AI partner (*M*=2.77, *SD*=1.32), but only in the Medium Similarity condition (Figure 4-3).

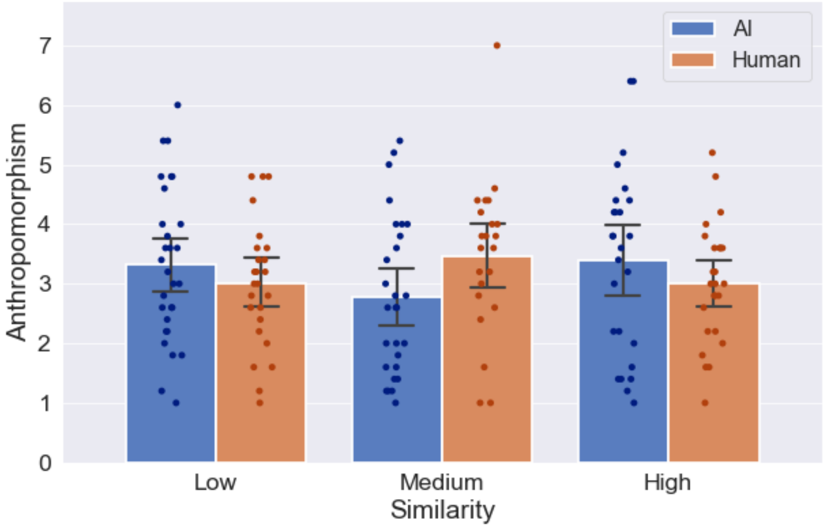


Figure 4-4. Bar plot for participants' ratings on their anthropomorphism towards their partner in Experiment 7. The dots represent the individual data points. Error bars represent 95% CI.

4.5.6 Liking and rapport scale

The Liking and Rapport Scale items showed a high scale reliability, *Cronbach's Alpha* =.922. We therefore calculated the aggregated mean across all the items for each participant. The ANOVA test on mean Liking and Rapport scores failed to find any significant main effects or interaction effect.

4.5.7 Conversation length

An ANOVA on the lengths of conversation between participants and Julie revealed a borderline main effect of Partner Identity, *F*(1, 148)= 3.38, *p*=.068, η^2=.022 (Figure 4-4). However, a closer look at pair-wise comparisons revealed that the difference was mostly driven by the significantly shorter conversations with AI partners than human partners in the *Medium Similarity* condition. Participants spent, on average, 2.17 minutes longer talking to a human partner (*M*=15.50, *SD*=0.87) than the AI partner (*M*=13.33, *SD*=1.33) in the Medium Similarity condition. However, the overall interaction effect was not statistically significant, *F*(2, 148)= 2.06, *p*=.131, η^2=.027.

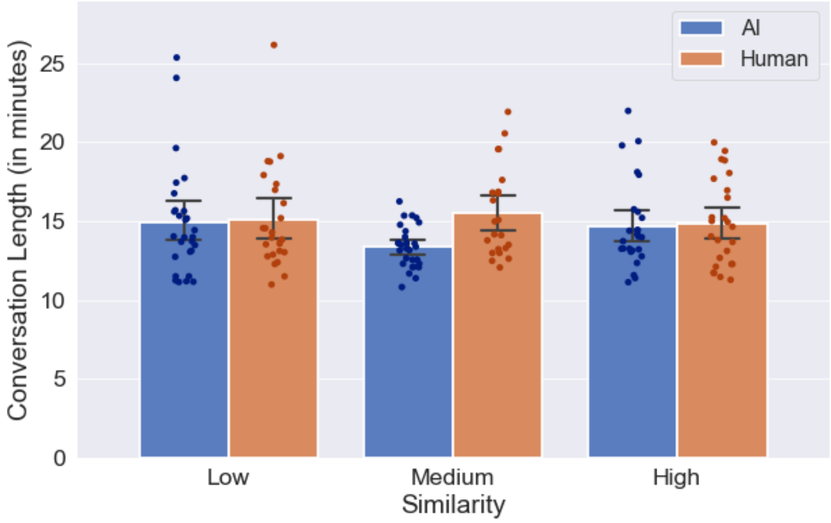


Figure 4-5. Bar plots for the amount of time (in minutes) that participants spent talking with their partner in Experiment 7. The dots represent the individual data points. Error bars represent 95% CI.

4.5.8 Facial action unit activities

ANOVAs on the mean AU4 activity, which is associated with negative expressions, during participants' conversation with Julie showed a significant Partner Identity main effect, *F*(1, 148)=8.82, *p*=.003, η^2=.056. Specifically, participants showed higher mean AU4 activity when they thought they were interacting with an AI (*M*=0.07, *SD*=.17) than a human partner (*M*=0.01, *SD*=.03). No other effects were found for AU4. ANOVAs on AU6 and AU12 mean activities, both of which are associated with positive expressions, did not yield any significant effects.

Bivariate correlations were conducted among all the dependent measures previously reported and with the AU activities. Most of the dependent measures were positively correlated with each other as well as with AU6 and AU12 activities. AU4, on the other hand, correlated with few of the dependent measures. See **Table 5** for the correlations among the dependent measures.

Table 6 Experiment 7 Correlations among dependent variables

	1	2	3	4	5	6	7	8	9	10
1. Warmth	-									
2. Competence	.674***	-								
3. MP-Agency	.666***	.695***	-							
4. MP-Experience	.455***	.425***	.720***	-						
5. Anthropomorphism	.595***	.515***	.523***	.411***	-					
6. Liking/Rapport	.731***	.594***	.578***	.398***	.673***	-				
7. Conversation Length	.251**	.114	.234**	.122	.160*	.228**	-			
8. AU4 activity	.093	.037	.040	.049	.115	.168*	.032	-		
9. AU6 activity	.222**	.083	.201*	.205*	.279***	.240**	.149	.261**	-	
10. AU12 activity	.223**	.097	.219**	.248**	.286***	.207*	.126	-.011	.851***	-

Note. **: significant at p=.05 level; **: significant at p=.01 level; ***: significant at p=.001 level*

4.6 Discussion

Overall, we did not find the proposed main effect of partner identity, nor did we find a main effect of self-disclosed similarity as the similarity-attraction effect predicts. However, a significant interaction between these two factors consistently emerged across almost all our dependent variables, except for the self-reported liking of the partner. It appeared that although our manipulations were robust enough to influence people's perceptions of their partner to some extent, they did not seem to adequately influence participants' liking of the partner, which is not uncommon in human-AI interaction studies (e.g., Bernier & Scassellati, 2010).

The significant interaction effect found across the dependent measures suggested that the uncanny valley effect, rather than the similarity-attraction effect, was a more likely underlying mechanism for our findings. Interestingly, the pattern of the interaction was not what we had expected. Specifically, the human partner was perceived more positively than the AI partner only when the partner self-disclosed a *medium* level of similarity to participants. Although unpredicted, this pattern of results is likely consistent with the uncanny valley effect when considering the most human-like self-disclosure behaviour in this experimental setting. When interacting with another human for the first time, an unusually high level of self-disclosed similarity (e.g., 12 out of 16 questions in the high similarity condition) from the other person might seem strange or even deliberative and manipulative, leading to negative perceptions of the partner. This might be especially true when the partner always responses after knowing one's own answer, as was the case in our experiment. In comparison, self-disclosing a medium level of similarity might appear as most natural and genuine, thus eliciting more favorable perceptions. This speculation was indirectly supported by the higher anthropomorphism rating for the human partner in the medium than the high and low similarity conditions (Figure 4-3). In other words, the human partner who self-disclosed a moderate amount of similarity to participants felt more human-like, potentially because the partner's disclosing behaviour was most natural and human-like given the specific context. In contrast, when participants talked to the AI partner who self-disclosed moderate similarity to themselves, the highly human-like style of self-disclosing behaviour might have induced the uncanny valley effect and motivated participants to perceive the AI especially negatively.

Moreover, participants’ self-reported anthropomorphism of their partners showed a similar pattern as their self-reported perceptions of their partner. Their anthropomorphism scores were also significantly and positively correlated with their scores on the perception measures. These findings provide support to our earlier argument that people’s anthropomorphism towards a partner might be influenced by social-cognitive processes, and is consistent with the perceptions of the partner, in addition to the anthropomorphic features.

The self-report measures were not only significantly correlated with each other with moderate to strong effect sizes (*r* =.398 to .731) but also mostly correlated with the conversation length, AU 6 activity, and AU 12 activity. Considering that AU6 and AU12 are indicators of smiling and positively correlate with self-reported happiness (Ekman et al., 1980), the overall positive correlations provide additional support for the validity of our dependent measures. AU4 activity was uncorrelated with most of our dependent measures and participants in the AI partner conditions showed greater such activity. Chatting with an AI is a unusual and ambiguous experience, which might have led to greater AU4 activity due to increased confusion, concentration, or novelty (Craig et al., 2008; McDaniel et al., 2007; Rozin & Cohen, 2003). Greater AU4 activity might thus reflect characteristics of the social context rather than participants’ impressions of or experiences with a conversation partner.

There are a couple limitations worth noting in the current study. First, even though we suspected that self-disclosing a moderate amount of similarity is the most natural human response in this specific social setting, which was indirectly supported by participants’ ratings on anthropomorphism towards their partner, this assumption was not directly tested. Replications are therefore needed to establish additional evidence for this account and future studies should directly assess whether people indeed perceive such responses as more natural and why. Second, the virtual human's limited body movement and facial expressions might have limited participants' engagement in the interactions, especially when participants believed that their partner was a human, and thus limited the effectiveness of our manipulation. Third, our participants recruited via Craigslist were demographically different from those we recruited on the university campus. Differences in demographic backgrounds might suggest different levels of exposure to technologies and thus different expectations of an AI partner, which might have led to different responses in our study. However, due to the limited number of participants we recruited via Craigslist (N=47 after exclusions), we did not have enough statistical power to examine potential group differences in addition to the effects of our manipulations.

Chapter 5

5 General Discussion

5.1 Does self-disclosure depth really matter in developing interpersonal liking between strangers?

We set out to answer the question of whether deep self-disclosures help induce greater liking of the discloser during initial interactions. The cumulative evidence obtained through our experiments suggested that they likely do not. Collins and Miller (1994) proposed that deeper self-disclosures would lead to greater liking of the discloser because (1) they communicate the discloser's liking and friendship intention and (2) the receiver might perceive the discloser to be more trusting and friendly, which leads to greater liking of them. In our experiments, self-disclosure depth did not influence participants' perceived friendship intention from the discloser (Experiment 3-4) or perceived trustworthiness or friendliness of the discloser (Experiment 1-4).

Interestingly, we observed some evidence suggesting that reciprocating a social partner's self-disclosure depth might lead to more positive outcomes in Experiment 5. Specifically, social partners who reciprocated the participants' own self-disclosure depth were considered as more responsive. Even though the effect of reciprocity on social preferences did not reach statistical significance in our experiment, the transactional model of intimacy (Reis & Shaver, 1988) would suggest that the higher perceived responsiveness of the partner is crucial to the development of a more intimate relationship. Moreover, participants seemed to "match" their reciprocity to their partner's level of reciprocity: if their partner rarely reciprocated their own self-disclosure depth, they reciprocated back less as well. These findings illustrated that people are highly capable of detecting subtleties in their social partner's behaviours and subsequently adjust their own behaviours based on such information.

In Experiment 6, we successfully replicated the effect that people self-disclose at deeper levels to those whom they like more, as suggested in Collins and Miller (1994), even after controlling for factors such as perceived knowledge and perceived similarity of the social partner. Taken together with findings from Experiments 1 through 5, it is more likely that increased self-disclosure depth is the result of interpersonal liking, rather than the cause of it. As real-life interactions involve dynamic and complicated interplay between self-disclosures, reciprocity of self-disclosures, and interpersonal liking, it is hard to disentangle the individual effects of these factors on each other. The previous studies that found deeper self-disclosures to lead to increased liking might have captured a fraction of the process, and thus provided a less accurate picture of the causal relationship between the two.

5.2 Similarity: The invisible hand of budding friendship

What, then, can lead to positive first impressions and initial liking if deep self-disclosures do not? The answer is unsurprising: similarity. Throughout our experiments, we consistently found a positive and robust *causal* effect of similarity on liking of a social partner whom participants had not previously met. Furthermore, our participants perceived social partners with higher similarity as showing greater interest in being their friend, as well as being more trustworthy and friendly, which might partially underlie their preferences for these partners. Given the strong empirical evidence in support of the similarity-attraction effect in the existing literature (e.g., Montoya & Horton, 2013; Montoya, Horton, & Kirchner, 2008; Tidwell, Eastwick, & Finkel, 2013), it was not at all surprising that we found such strong effects of similarity.

Our findings thus seem to support our suspicion that similarity between the social partners is an “invisible hand” behind the disclosure-liking relationship. Similarity likely contributes to increased mutual liking and comfort with another, which in turn leads to greater exchange of deeper self-disclosures, resulting in positive feelings and enhanced relationship potential. We also proposed another possibility in which similarity might confound the disclosure-liking relationship at the beginning: a high level of similarity might enable the effect of deep self-disclosures on liking, because the receiver might welcome deep disclosures more from a similar than a dissimilar other. However, the fact that we failed to find any significant interaction effect between similarity and disclosure depth throughout our experiments suggested against this possibility. If similarity was the enabling factor for disclosure depth, we would have observed an increased liking of the deeper discloser when the similarity between the two partners was high.

One important and interesting caveat to point out is that how such similarity is communicated also seems to play an important role in shaping liking of the discloser. In our first six experiments, the self-disclosed similarity was communicated in a less “contrived” way in that it was unlikely that participants believed that the others “crafted” their information to match their own. In Experiment 7, however, where participants always self-disclosed before the virtual human responded to the same question, participants might have perceived a high level of self-disclosed similarity from the partner as unnatural at best and ingenuine and manipulative at worst. Consequently, we failed to find the linear relationship between similarity and liking of the partner. Rather, participants seemed to form most favorable impressions of the human partner when the partner self-disclosed a medium level of similarity, which might have been perceived as more genuine. This speculation is consistent with the reinforcement-affect account of interpersonal attraction (e.g., Byrne & Nelson, 1965; Clore & Byrne, 1974), which asserts that people’s attraction to a social partner is a function of their positive affective responses, in relation to the negative affective responses, associated with that person. It seems that while similarities often elicit positive affective responses, excessive expressions of similarities, especially when perceived as contrived, can elicit negative affective responses to the discloser, which might outweigh the benefit of the similarities and lead to decreased liking of the discloser.

5.3 Limitations

Although the specific limitations were discussed in each of the empirical chapters, we would like to point out two general shortcomings of our experiments that limit the scope of the work. First, a big strength of the current work is the careful experimental control, which allowed us to disentangle complicated effects and reduce the noise of potential confounds. However, this meticulous control also created the biggest weakness of the research— the limited ecological validity and reduced generalizability to real-world interactions. Self-disclosure is not a one-shot, static event that occurs in isolation. Rather, it is such a complicated and dynamic process in which various situational and dispositional factors interplay to influence people's decisions to make disclosures, their responses to others' disclosures, and how they flexibly adjust their responses and behaviours minute-by-minute. Hence, while our experiments provided insights into some contributing factors and underlying mechanisms of the process, they were far from being able to provide a "big picture" that accurately describes how this process unfolds in real life interactions.

Second, the student samples used throughout our experiments further limit the generalizability of our findings. Due to the nature of the experiments and limited resources, we were not able to recruit participants with more diverse demographic backgrounds. Our findings might thus reflect characteristics of this specific demographic group. The university student samples were largely WEIRD samples, consisting of people from Western, educated, industrialized, rich, and democratic cultural backgrounds, who are "frequent outliers" in various psychological characteristics compared to other populations (Henrich et al., 2010; Jones, 2010). Future studies that include more diverse samples are needed before we can conclude that our findings represent general social interaction processes instead of patterns specifically found in university students and WEIRD populations. Moreover, as the student samples are relatively homogeneous, our findings might have reflected interaction patterns between people who are already relatively similar to each other rather than those who are not.

5.4 Conclusion

Despite these limitations, we made a few contributions to the literature with the current line of work. First, we provided strong evidence that shows, in contrast to what was suggested in the literature, self-disclosure depth per se does not cause interpersonal liking in initial interactions. Rather, the causal relationship is more likely to be the other way around. Second, we identified similarity between social partners as a key factor that underlies the disclosure-liking relationship: greater similarity promotes liking, which in turn leads to greater exchange of deeper self-disclosures. Finally, we tried to conceptually replicate a few effects identified in the existing literature: the effect of self-disclosure depth on liking, the effect of liking on self-disclosure, and the similarity-attraction effect. Two of these effects were successfully replicated and one was not. In the face of the issues surrounding replicability of psychological research, we tested these effects to our best abilities using pre-registrations, repeated testing, and meta-analysis (Maxwell et al., 2015). Falsifiability lies at the heart of science. We provided cumulative evidence that contributes to the falsification process, which is a key to establishing sound psychological science. Thus, our final conclusion is that while disclosure depth plays an important role in relationship development over longer timescales, similarity seems to be the deciding factor in whether people choose to initiate a relationship in the first place.

References

Abram, S. V, Breton, Y.-A., Schmidt, B., Redish, A. D., & MacDonald, A. W. (2016). The Web-Surf Task: A translational model of human decision-making. *Cognitive, Affective & Behavioral Neuroscience*, *16*(1), 37–50. https://doi.org/10.3758/s13415-015-0379-y

Ajzen, I. (1974). Effects of information on interpersonal attraction: Similarity versus affective value. *Journal of Personality and Social Psychology*, *29*(3), 374–380. https://doi.org/10.1037/h0036002

Ajzen, I. (1977). Information Processing Approaches to Interpersonal Attraction. *Theory and Practice in Interpseronal Attraction*, *January 1977*, 51–77.

Alter, A. L., & Oppenheimer, D. M. (2009a). Uniting the Tribes of Fluency to Form a Metacognitive Nation. *Personality and Social Psychology Review*, *13*(3), 219–235. https://doi.org/10.1177/1088868309341564

Alter, A. L., & Oppenheimer, D. M. (2009b). Suppressing Secrecy Through Metacognitive Ease: Cognitive Fluency Encourages Self-Disclosure. *Psychological Science*, *20*(11), 1414–1420. https://doi.org/10.1111/j.1467-9280.2009.02461.x

Altman, I., & Taylor, D. A. (Dalmas A. (1973). *Social penetration: the development of interpersonal relationships*. Holt, Rinehart and Winston. http://psycnet.apa.org/psycinfo/1973-28661-000

Archer, R. L., & Berg, J. H. (1978). Disclosure reciprocity and its limits: A reactance analysis. *Journal of Experimental Social Psychology*, *14*(6), 527–540. https://doi.org/10.1016/0022-1031(78)90047-1

Archer, R. L., Hormuth, S. E., & Berg, J. H. (1982). Avoidance of Self-Disclosure. *Personality and Social Psychology Bulletin*, *8*(1), 122–128. https://doi.org/10.1177/014616728281019

Arend, M. G., & Schäfer, T. (2019). Statistical power in two-level models: A tutorial based on monte carlo simulation. *Psychological Methods*, *24*(1), 1–19. https://doi.org/10.1037/met0000195

Aronson, E., & Worchel, P. (1966). Similarity versus liking as determinants of interpersonal attractiveness. *Psychonomic Science*, *5*(4), 157–158. https://doi.org/10.3758/BF03328329

Artstein, R., Traum, D., Boberg, J., Gainer, A., Gratch, J., Johnson, E., Leuski, A., & Nakano, M. (2016). Niki and Julie: a robot and virtual human for studying multimodal social interaction. *Proceedings of the 18th ACM International Conference on Multimodal Interaction - ICMI 2016*, 402–403. https://doi.org/10.1145/2993148.2998532

Atmaca, S., Sebanz, N., & Knoblich, G. (2011). The joint flanker effect: Sharing tasks with real and imagined co-actors. *Experimental Brain Research*, *211*(3–4), 371–385. https://doi.org/10.1007/s00221-011-2709-9

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, *6*(2), 269–278. https://doi.org/10.1037/1528-3542.6.2.269

Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2007). Is The Uncanny Valley An Uncanny Cliff? *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*, 368–373. https://doi.org/10.1109/ROMAN.2007.4415111

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics*, *1*(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v067.i01

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*(3), 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Bauminger, N., Finzi-Dottan, R., Chason, S., & Har-Even, D. (2008). Intimacy in adolescent friendship: The roles of attachment, coherence, and self-disclosure. *Journal of Social and Personal Relationships*, *25*(3), 409–428. https://doi.org/10.1177/0265407508090866

Beck, I., Smits, D. J. M., Claes, L., Vandereycken, W., & Bijttebier, P. (2009). Psychometric evaluation of the behavioral inhibition/behavioral activation system scales and the sensitivity to punishment and sensitivity to reward questionnaire in a sample of eating disordered patients. *Personality and Individual Differences*, *47*(5), 407–412. https://doi.org/10.1016/j.paid.2009.04.007

Berger, C., & Calabrese, R. (1975). Some exprolations in initial interaction and beyond: Toward a development theory of interpersonal communication. *Human Communication Research*, *1*, 99–112.

Bernier, E. P., & Scassellati, B. (2010). The similarity-attraction effect in human-robot interaction. *2010 IEEE 9th International Conference on Development and Learning*, 286–290. https://doi.org/10.1109/DEVLRN.2010.5578828

Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica*, *22*(1), 23. https://doi.org/10.2307/1909829

Bickmore, T., Schulman, D., & Yin, L. (2009). Engagement vs. Deceit: Virtual Humans with Human Autobiographies. In Z. Ruttkay, M. Kipp, A. Nijholt, & H. H. Vilhjálmsson (Eds.), *Lecture Notes in Computer Science* (Vol. 5773, pp. 6–19). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04380-2_4

Boone, T. R., & Buck, R. (2003). Emotional expressivity and trustworthiness: The role of nonverbal behavior in the evolution of cooperation. *Journal of Nonverbal Behavior*, *27*(3), 163–182. https://doi.org/10.1023/A:1025341931128

Bornovalova, M. A., Cashman-Rolls, A., O'Donnell, J. M., Ettinger, K., Richards, J. B., DeWit, H., & Lejuez, C. W. (2009). Risk taking differences on a behavioral task as a function of potential reward/loss magnitude and individual differences in impulsivity and sensation seeking. *Pharmacology Biochemistry and Behavior*, *93*(3), 258–262. https://doi.org/10.1016/j.pbb.2008.10.023

Braddock, K. H., Dillard, J. P., Voigt, D. C., Stephenson, M. T., Sopory, P., & Anderson, J. W. (2011). Impulsivity Partially Mediates the Relationship Between BIS/BAS and Risky Health Behaviors. *Journal of Personality*, *79*(4), 793–810. https://doi.org/10.1111/j.1467-6494.2011.00699.x

Brennan, K. A., Clark, C. L., & Shaver, P. R. (1998). Self-report measurement of adult attachment: An integrative overview. In *Attachment theory and close relationships*.

Buelow, M. T., & Blaine, A. L. (2015). The assessment of risky decision making: A factor analysis of performance on the Iowa Gambling Task, Balloon Analogue Risk Task, and Columbia Card Task. *Psychological Assessment*, *27*(3), 777–785. https://doi.org/10.1037/a0038622

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Byrne, D. (1971). *The attraction paradigm*. Academic Press.

Byrne, D. (1997). An overview (and underview) of research and theory within the attraction paradigm. In *Journal of Social and Personal Relationships*. https://doi.org/10.1177/0265407597143008

Byrne, D., & Clore, G. L. (1967). Effectance arousal and attraction. *Journal of Personality and Social Psychology*, *6*(4, Pt.2), 1–18. https://doi.org/10.1037/h0024829

Byrne, D., & Nelson, D. (1965). Attraction as a linear function of proportion of positive reinforcements. *Journal of Personality and Social Psychology*, *1*(6), 659–663. https://doi.org/10.1037/h0022073

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, *125*(3), 367–383. https://doi.org/10.1037/0033-2909.125.3.367

Caltabiano, M. L., & Smithson, M. (1983). Variables affecting the perception of self-disclosure appropriateness. *Journal of Social Psychology*, *120*(1), 119–128. https://doi.org/10.1080/00224545.1983.9712017

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology*, *67*(2), 319–333. https://doi.org/10.1037/0022-3514.67.2.319

Chaikin, A. L., & Derlega, V. J. (1974). Variables affecting the appropriateness of self-disclosure. *Journal of Consulting and Clinical Psychology*, *42*(4), 588–593. https://doi.org/10.1037/h0036614

Christensen, R. H. B. (2015). *ordinal—regression models for ordinal data* (R package version 28).

Christensen, R. H. B. (2019). Cumulative Link Models for Ordinal Regression with the R Package ordinal. *Journal of Statistical Software*.

Clore, G. L., & Byrne, D. (1974). A Reinforcement-Affect Model of Attraction. In *Foundations of Interpersonal Attraction*. https://doi.org/10.1016/b978-0-12-362950-0.50013-6

Cohen, S. (2004). Social Relationships and Health. *American Psychologist*, *59*(8), 676–684. https://doi.org/10.1037/0003-066X.59.8.676

Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin*, *116*(3), 457–475. https://doi.org/10.1037/0033-2909.116.3.457

Condon, J. W., & Crano, W. D. (1988). Inferred evaluation and the relation between attitude similarity and interpersonal attraction. *Journal of Personality and Social Psychology*, *54*(5), 789–797. https://doi.org/10.1037/0022-3514.54.5.789

Coyne, J. C. (1976). Depression and the response of others. *Journal of Abnormal Psychology*, *85*(2), 186–193. https://doi.org/10.1037/0021-843X.85.2.186

Craig, S. D., D'Mello, S., Witherspoon, A., & Graesser, A. (2008). Emote aloud during learning with AutoTutor: Applying the Facial Action Coding System to cognitive–affective states during learning. *Cognition & Emotion*, *22*(5), 777–788. https://doi.org/10.1080/02699930701516759

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, *43*(3), 245–265. https://doi.org/10.1348/0144665031752934

Cropanzano, R., & Mitchell, M. S. (2005). Social Exchange Theory: An Interdisciplinary Review. *Journal of Management*, *31*(6), 874–900. https://doi.org/10.1177/0149206305279602

Cunningham, M. R. (1988). Does Happiness Mean Friendliness? *Personality and Social Psychology Bulletin*, *14*(2), 283–297. https://doi.org/10.1177/0146167288142007

de Melo, C. M., Carnevale, P. J., & Gratch, J. (2014). Social Categorization and Cooperation between Humans and Computers. *Proceedings of the Annual Meeting of the Cognitive Science Society.*, *36*(36).

Deakin, J., Aitken, M., Robbins, T., & Sahakian, B. J. (2004). Risk taking during decision-making in normal volunteers changes with age. *Journal of the International Neuropsychological Society*. https://doi.org/10.1017/S1355617704104104

Derlega, V. J. (2009). Social Penetration Theory. In *Encyclopedia of Human Relationships*. SAGE Publications, Inc. https://doi.org/10.4135/9781412958479.n511

Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin, 81*(6), 358–361. https://doi.org/10.1037/h0036316

Ekman, P., Freisen, W. V., & Ancoli, S. (1980). Facial signs of emotional experience. *Journal of Personality and Social Psychology, 39*(6), 1125–1134. https://doi.org/10.1037/h0077722

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124–129. https://doi.org/10.1037/h0030377

Ekman, P., & Rosenberg, E. L. (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS).* Oxford University Press, USA.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Fernie, G., Cole, J. C., Goudie, A. J., & Field, M. (2010). Risk-taking but not response inhibition or delay discounting predict alcohol consumption in social drinkers. *Drug and Alcohol Dependence*, *112*(1–2), 54–61. https://doi.org/10.1016/j.drugalcdep.2010.05.011

Ferrari, F., Paladino, M. P., & Jetten, J. (2016). Blurring Human–Machine Distinctions: Anthropomorphic Appearance in Social Robots as a Threat to Human Distinctiveness. *International Journal of Social Robotics*, *8*(2), 287–302. https://doi.org/10.1007/s12369-016-0338-y

Figner, B., Mackinlay, R. J., Wilkening, F., & Weber, E. U. (2009). Affective and Deliberative Processes in Risky Choice: Age Differences in Risk Taking in the Columbia Card Task. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*(3), 709–730. https://doi.org/10.1037/a0014983

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Forgas, J. P. (2011). Affective influences on self-disclosure: Mood effects on the intimacy and reciprocity of disclosing personal information. *Journal of Personality and Social Psychology*, *100*(3), 449–461. https://doi.org/10.1037/a0021129

Fowles, D. C. (1980). The Three Arousal Model: Implications of Gray's Two-Factor Learning Theory for Heart Rate, Electrodermal Activity, and Psychopathy. *Psychophysiology*, *17*(2), 87–104. https://doi.org/10.1111/j.1469-8986.1980.tb00117.x

Gaher, R. M., Hahn, A. M., Shishido, H., Simons, J. S., & Gaster, S. (2015). Associations between sensitivity to punishment, sensitivity to reward, and gambling. *Addictive Behaviors*, *42*, 180–184. https://doi.org/10.1016/j.addbeh.2014.11.014

Gouldner, A. W. (1960). The Norm of Reciprocity : A Preliminary Statement. *American Sociological Review*, *25*(2), 161–178. https://doi.org/10.2307/2092623

Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. (2009). Creating Rapport with Virtual Agents. In Zsófia Ruttkay, M. Kipp, A. Nijholt, & H. H. Vilhjálmsson (Eds.), *Intelligent Virtual Agents* (Vol. 5773, Issue May 2014, pp. 125–138).

Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-74997-4_12

Gray, H. M., Gray, K., & Wegner, D. M. (2014). Dimensions of Mind Perception. In *Experimental Philosophy* (Vol. 315, Issue 5812, pp. 77–80). Oxford University Press. https://doi.org/10.1126/science.1134475

Gray, J. A. (1981). A Critique of Eysenck's Theory of Personality. In *A Model for Personality* (pp. 246–276). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-67783-0_8

Gray, J. A. (1982). Précis of The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system. *Behavioral and Brain Sciences*, *5*(3), 469–484. https://doi.org/10.1017/S0140525X00013066

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, *125*(1), 125–130. https://doi.org/10.1016/j.cognition.2012.06.007

Harless, D. W., & Camerer, C. F. (1994). The Predictive Utility of Generalized Expected Utility Theories. *Econometrica*, *62*(6), 1251. https://doi.org/10.2307/2951749

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review*, *10*(3), 252–264. https://doi.org/10.1207/s15327957pspr1003_4

Hegel, F., Krach, S., Kircher, T., Wrede, B., & Sagerer, G. (2008). Understanding social robots: A user study on anthropomorphism. *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, 574–579. https://doi.org/10.1109/ROMAN.2008.4600728

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Holman, L., Head, M. L., Lanfear, R., & Jennions, M. D. (2015). Evidence of experimental bias in the life sciences: Why we need blind data recording. *PLoS Biology*, *13*(7). https://doi.org/10.1371/journal.pbio.1002190

Hopko, D. R., Lejuez, C. W., Daughters, S. B., Aklin, W. M., Osborne, A., Simmons, B. L., & Strong, D. R. (2006). Construct Validity of the Balloon Analogue Risk Task (BART): Relationship with MDMA Use by Inner-City Drug Users in Residential

Treatment. *Journal of Psychopathology and Behavioral Assessment*, *28*(2). https://doi.org/10.1007/s10862-006-7487-5

Hox, J. J. (2010). Multilevel analysis: Techniques and applications: Second edition. In *Multilevel Analysis: Techniques and Applications: Second Edition*. https://doi.org/10.4324/9780203852279

Human, L. J., Sandstrom, G. M., Biesanz, J. C., & Dunn, E. W. (2013). Accurate First Impressions Leave a Lasting Impression: The Long-Term Effects of Distinctive Self-Other Agreement on Relationship Development. *Social Psychological and Personality Science*, *4*(4), 395–402. https://doi.org/10.1177/1948550612463735

iMotions. (2016). *Facial Expression Analysis: The Complete Pocket Guide*. https://imotions.com/guides/facial-expression-analysis/

Insko, C. A., & et al. (1973). Implied evaluation and the similarity-attraction effect. *Journal of Personality and Social Psychology*, *25*(3), 297–308. https://doi.org/10.1037/h0034224

Jones, D. (2010). A WEIRD View of Human Nature Skews Psychologists' Studies. *Science*, *328*(5986), 1627–1627. https://doi.org/10.1126/science.328.5986.1627

Jourard, S. M. (1959). Self-disclosure and other-cathexis. *The Journal of Abnormal and Social Psychology*, *59*(3), 428–431. https://doi.org/10.1037/h0041640

Kang, S. H., & Gratch, J. (2011). People like virtual counselors that highly-disclose about themselves. *Annual Review of CyberTherapy and Telemedicine*, *9*(1), 117–120.

Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology*, *6*(MAR), 1–16. https://doi.org/10.3389/fpsyg.2015.00390

Keelan, J. P. R., Dion, K. K., & Dion, K. L. (1998). Attachment style and relationship satisfaction: Test of a self-disclosure explanation. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement*, *30*(1), 24–35. https://doi.org/10.1037/h0087055

Kerr, M., Kan, H. Ê., And, S., & Trost, K. (1999). *To know you is to trust you: parents' trust is rooted in child disclosure of information*. https://ac.elscdn.com/S014019719990266X/1-s2.0-S014019719990266X-main.pdf?_tid=dc886cda-bae7-422e-ba4a-d736102ea0e3&acdnat=1529344224_fe1b9c1348523e2a97c6501ed8a0d256

Kim, D.-Y., & Lee, J.-H. (2011). Effects of the BAS and BIS on decision-making in a gambling task. *Personality and Individual Differences*, *50*(7), 1131–1135. https://doi.org/10.1016/j.paid.2011.01.041

Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin and Review*, *20*(1), 54–72. https://doi.org/10.3758/s13423-012-0341-8

Lauriola, M., Panno, A., Levin, I. P., & Lejuez, C. W. (2014). Individual Differences in Risky Decision Making: A Meta-analysis of Sensation Seeking and Impulsivity with the Balloon Analogue Risk Task. *Journal of Behavioral Decision Making*, *27*(1), 20–36. https://doi.org/10.1002/bdm.1784

Laursen, B., & Hartup, W. W. (2002). The Origins of Reciprocity and Social Exchange in Friendships. *New Directions for Child and Adolescent Development*, *2002*(95), 27–40. https://doi.org/10.1002/cd.35

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology. Applied*, *8*(2), 75–84. https://doi.org/10.1037//1076-898x.8.2.75

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, *127*(2), 267–286. https://doi.org/10.1037//0033-2909.127.2.267

Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. P. (2014). It’s only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94–100. https://doi.org/10.1016/j.chb.2014.04.043

Lydon, J. E., Jamieson, D. W., & Zanna, M. P. (1988). Interpersonal Similarity and the Social and Intellectual Dimensions of First Impressions. *Social Cognition*, *6*(4), 269–286. https://doi.org/10.1521/soco.1988.6.4.269

Maas, C. J. M., & Hox, J. J. (2005). Sufficient Sample Sizes for Multilevel Modeling.

Methodology, *1*(3), 86–92. https://doi.org/10.1027/1614-2241.1.3.86

MacDorman, K. F., & Chattopadhyay, D. (2016). Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition*, *146*, 190–205. https://doi.org/10.1016/j.cognition.2015.09.019

MacDorman, K. F. K. (2006). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. *ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science, January 2006*, 26–29. https://doi.org/10.1093/scan/nsr025

Manohar, S. G., & Husain, M. (2013). Attention as foraging for information and value. *Frontiers in Human Neuroscience*, *7*(November), 711. https://doi.org/10.3389/fnhum.2013.00711

Marek, C. I., Wanzer, M. B., & Knapp, J. L. (2004). An exploratory investigation of the relationship between roommates' first impressions and subsequent communication patterns. *Communication Research Reports*, *21*(2), 210–220. https://doi.org/10.1080/08824090409359982

Markiewicz, Ł., & Kubińska, E. (2015). Information Use Differences in Hot and Cold Risk Processing: When Does Information About Probability Count in the Columbia Card Task? *Frontiers in Psychology*, *6*(NOV), 1727. https://doi.org/10.3389/fpsyg.2015.01727

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist*, *70*(6), 487–498. https://doi.org/10.1037/a0039400

McCarty, J. A., & Shrum, L. J. (2000). The Measurement of Personal Values in Survey Research. *Public Opinion Quarterly*, *64*(3), 271–298. https://doi.org/10.1086/317989

McDaniel, B., D'Mello, S., King, B., Chipman, P., Tapp, K., & Graesser, A. (2007). *Facial Features for Affective State Detection in Learning Environments BT -*

Proceedings of the 29th Annual Meeting of the Cognitive Science Society.

McGinley, H., McGinley, P., & Nicholas, K. (1978). Smiling, body position, and interpersonal attraction. *Bulletin of the Psychonomic Society*, *12*(1), 21–24. https://doi.org/10.3758/BF03329613

Mikulincer, M., & Nachshon, O. (1991). Attachment Styles and Patterns of Self-Disclosure. *Journal of Personality and Social Psychology*, *61*(2), 321–331. https://doi.org/10.1037/0022-3514.61.2.321

Mikulincer, M., Shaver, P. R., & Pereg, D. (2003). Attachment Theory and Affect Regulation: The Dynamics, Development, and Cognitive Consequences of Attachment-Related Strategies. *Motivation and Emotion*, *27*(2), 77–102. https://link.springer.com/content/pdf/10.1023%2FA%3A1024515519160.pdf

Mills, B., Reyna, V. F., & Estrada, S. (2008). Explaining Contradictory Relations Between Risk Perception and Risk Taking. *Psychological Science*, *19*(5), 429–433. https://doi.org/10.1111/j.1467-9280.2008.02104.x

Mitchell, W. J., Szerszen, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & MacDorman, K. F. (2011). A Mismatch in the Human Realism of Face and Voice Produces an Uncanny Valley. *I-Perception*, *2*(1), 10–12. https://doi.org/10.1068/i0415

Montoya, R. M., & Horton, R. S. (2013). A meta-analytic investigation of the processes underlying the similarity-attraction effect. *Journal of Social and Personal Relationships*, *30*(1), 64–94. https://doi.org/10.1177/0265407512452989

Montoya, R. M., Horton, R. S., & Kirchner, J. (2008). Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships*, *25*(6), 889–922. https://doi.org/10.1177/0265407508096700

Moon, Y. (2000). Intimate Exchanges: Using Computers to Elicit Self-Disclosure From Consumers. *Journal of Consumer Research*, *26*(4), 323–339. https://doi.org/10.1086/209566

Moon, Y., & Nass, C. (1996). How “Real” Are Computer Personalities? *Communication Research*, *23*(6), 651–674. https://doi.org/10.1177/009365096023006002

Moreland, R. L., & Zajonc, R. B. (1982). Exposure effects in person perception:

Familiarity, similarity, and attraction. *Journal of Experimental Social Psychology*, *18*(5), 395–415. https://doi.org/10.1016/0022-1031(82)90062-2

Mori, M. (1970). The Uncanny Valley. *Energy*, *7*(4), 33–35.

Mubin, O., Stevens, C. J., Shahid, S., Mahmud, A. Al, & Dong, J.-J. (2013). A Review of the Applicability of Robots in Education. *Technology for Education and Learning*, *1*(1). https://doi.org/10.2316/Journal.209.2013.1.209-0015

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

O'Connor, R. M., Colder, C. R., & Hawk, L. W. (2004). Confirmatory factor analysis of the Sensitivity to Punishment and Sensitivity to Reward Questionnaire. *Personality and Individual Differences*, *37*(5), 985–1002. https://doi.org/10.1016/j.paid.2003.11.008

O'Connor, R. M., Stewart, S. H., & Watt, M. C. (2009). Distinguishing BAS risk for university students' drinking, smoking, and gambling behaviors. *Personality and Individual Differences*, *46*(4), 514–519. https://doi.org/10.1016/j.paid.2008.12.002

Osborne, J. W., Costello, A. B., & Kellow, J. T. (2008). Best practices in exploratory factor analysis. *Best Practices in Quantitative Methods*, 86–99. https://methods.sagepub.com/base/download/BookChapter/best-practices-in-quantitative-methods/d8.xml

Ostir, G. V., Smith, P. M., Smith, D., & Ottenbacher, K. J. (2005). Reliability of the positive and negative affect schedule (PANAS) in medical rehabilitation. *Clinical Rehabilitation*, *19*(7), 767–769. https://doi.org/10.1191/0269215505cr894oa

Pennebaker, J. W. (1985). Traumatic experience and psychosomatic disease: Exploring the roles of behavioural inhibition, obsession, and confiding. *Canadian Psychology/Psychologie Canadienne*, *26*(2), 82–95. https://doi.org/10.1037/h0080025

Penolazzi, B., Gremigni, P., & Russo, P. M. (2012). *Impulsivity and Reward Sensitivity differentially influence affective and deliberative risky decision making*. https://doi.org/10.1016/j.paid.2012.05.018

Rao, H., Korczykowski, M., Pluta, J., Hoang, A., & Detre, J. A. (2008). Neural correlates of voluntary and involuntary risk taking in the human brain: An fMRI Study of the Balloon Analog Risk Task (BART). *NeuroImage*, *42*(2), 902–910. https://doi.org/10.1016/j.neuroimage.2008.05.046

Reis, H. T., Maniaci, M. R., Caprariello, P. A., Eastwick, P. W., & Finkel, E. J. (2011). Familiarity does indeed promote attraction in live interaction. *Journal of Personality and Social Psychology*, *101*(3), 557–570. https://doi.org/10.1037/a0022885

Reis, H. T., & Shaver, P. (1988). Intimacy as an interpersonal process. In S. Duck, D. F. Hay, S. E. Hobfoll, W. Ickes, & B. M. Montgomery (Eds.), *Handbook of personal relationships: Theory, research and interventions* (pp. 367–389). John Wiley & Sons, Ltd.

Robinson, H., MacDonald, B., & Broadbent, E. (2014). The Role of Healthcare Robots for Older People at Home: A Review. *International Journal of Social Robotics*, *6*(4), 575–591. https://doi.org/10.1007/s12369-014-0242-2

Rozin, P., & Cohen, A. B. (2003). High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans. *Emotion*, *3*(1), 68–75. https://doi.org/10.1037/1528-3542.3.1.68

Ryb, G. E., Dischinger, P. C., Kufera, J. A., & Read, K. M. (2006). Risk perception and impulsivity: Association with risky behaviors and substance abuse disorders. *Accident Analysis and Prevention*, *38*(3), 567–573. https://doi.org/10.1016/j.aap.2005.12.001

Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, *69*(4), 719–727. https://doi.org/10.1037/0022-3514.69.4.719

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience*, *7*(4), 413–422. https://doi.org/10.1093/scan/nsr025

Scherwitz, L., & Helmreich, R. (1973). Interactive effects of eye contact and verbal content on interpersonal attraction in dyads. *Journal of Personality and Social Psychology*, *25*(1), 6–14. https://doi.org/10.1037/h0034270

Seyama, J., & Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments*, *16*(4), 337–351. https://doi.org/10.1162/pres.16.4.337

Slovic, P., Peters, E., Finucane, M. L., & MacGregor, D. G. (2005). Affect, risk, and decision making. *Health Psychology*, *24*(4, Suppl), S35–S40. https://doi.org/10.1037/0278-6133.24.4.S35

Stein, J. P., & Ohler, P. (2017). Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition*, *160*, 43–50. https://doi.org/10.1016/j.cognition.2016.12.010

Sunnafrank, M., & Ramirez, A. (2004). At first sight: Persistent relational effects of get-acquainted conversations. *Journal of Social and Personal Relationships*, *21*(3), 361–379. https://doi.org/10.1177/0265407504042837

Taylor, D. A. (1979). Motivational bases. In G. J. Chelune (Ed.), *Self-disclosure: Origins, patterns, and implications of openness in interpersonal relationships* (pp. 110–151). Jossey-Bass.

Taylor, D. A., & Altman, I. (1975). Self-Disclosure as a Function of Reward-Cost Outcomes. *Sociometry*, *38*(1), 18. https://doi.org/10.2307/2786231

Thurstone, L. L. (1934). The vectors of mind. *Psychological Review*, *41*(1), 1–32. https://doi.org/10.1037/h0075959

Tidwell, N. D., Eastwick, P. W., & Finkel, E. J. (2013). Perceived, not actual, similarity predicts initial attraction in a live romantic context: Evidence from the speed-dating paradigm. *Personal Relationships*, *20*(2), 199–215. https://doi.org/10.1111/j.1475-6811.2012.01405.x

Tinwell, A., Grimshaw, M., & Nabi, D. A. (2015). The effect of onset asynchrony in audio-visual speech and the Uncanny Valley in virtual characters. *International Journal of Mechanisms and Robotic Systems*, *2*(2), 97. https://doi.org/10.1504/ijmrs.2015.068991

Tinwell, A., Grimshaw, M., Nabi, D. A., & Williams, A. (2011). Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Computers in Human Behavior*, *27*(2), 741–749. https://doi.org/10.1016/j.chb.2010.10.018

Torrubia, R., Ávila, C., Moltó, J., & Caseras, X. (2001). The Sensitivity to Punishment and Sensitivity to Reward Questionnaire (SPSRQ) as a measure of Gray's anxiety and impulsivity dimensions. *Personality and Individual Differences*, *31*(6), 837–862. https://doi.org/10.1016/S0191-8869(00)00183-5

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Verberne, F. M. F., Ham, J., & Midden, C. J. H. (2015). Trusting a Virtual Driver That Looks, Acts, and Thinks Like You. *Human Factors*, *57*(5), 895–909. https://doi.org/10.1177/0018720815580749

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wei, M., Russell, D. W., Mallinckrodt, B., & Vogel, D. L. (2007). The Experiences in Close Relationship Scale (ECR)-short form: Reliability, validity, and factor structure. *Journal of Personality Assessment*, *88*(2), 187–204. https://doi.org/10.1080/00223890701268041

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking. *Frontiers in Psychology*, *7*(NOV), 1832. https://doi.org/10.3389/fpsyg.2016.01832

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Zuckerman, M., & Kuhlman, D. M. (2000). Personality and risk-taking: Common biosocial factors. *Journal of Personality*, *68*(6), 999–1029.

https://doi.org/10.1111/1467-6494.00124

Appendices

Appendix A: Multiple-Choice Questions Used in the Experiments

Low Self-Disclosure Depth Questions:

Question 1: What's your favorite season?
1. The fresh air in spring makes me happy.
2. I love spending summer days outdoors.
3. I love jumping into the leaves in autumn.
4. Building snowmen is great winter fun.

Question 2: Which of these social media sites do you prefer?
1. Facebook makes keeping in touch with my friends really easy.
2. Instagram allows me to enjoy my friends' special moments.
3. Sending silly faces to my friends over snapchat is so much fun.
4. Hashtags on Twitter are hilarious.

Question 3: How heavy is your course workload this semester?
1. I'm only taking a couple of courses this semester, it's a breeze.
2. I'm taking a few courses this semester, but I've been able to manage everything.
3. I'm very busy with school work this semester.
4. I barely have time to sleep with all the schoolwork I have to do.

Question 4: Which of these movie genres do you prefer?
1. I like to imagine the happy endings in romance movies.
2. Comedy movies make light of things and always make me laugh.
3. I like how excited action movies make me feel.
4. I like how dark and twisted horror movies can be.

Question 5: Which of these sports do you like best?
1. I play soccer regularly and I enjoy it very much.
2. Some of my fondest childhood memories happened at hockey tournaments.
3. I love watching competitive swimming on TV.
4. I love the fast pace of basketball games.

Question 6: Which of these cuisines do you like best?
1. I could eat Italian food everyday, especially pasta and desserts.
2. Japanese food always looks pretty and tastes fresh.
3. I love that Chinese dishes are made to be shared with others.
4. I always get pretty excited about spicy Mexican dishes.

Question 7: Which of these holidays is your favourite?
1. I enjoy giving and receiving Christmas gifts very much.
2. I love celebrating New Year's Eve with my friends and family.
3. Halloween is the most fun holiday of the year.
4. I love turkey and pumpkin pie on Thanksgiving.

Question 8: Which of these overseas countries would you most like to visit?
1. Madrid and Barcelona are on the top of my list for travel destinations.
2. Australia's beautiful coastline is something I long to see.
3. I would really love to visit Venice and take a gondola ride along the canals.
4. Japan would be a lot of fun to visit because I love animation and cosplay.

Question 9: Which of these activities do you prefer for a vacation?
1. I love spending my vacations travelling around and experiencing new cultures.
2. My favourite vacation is to stay at home and just relax.
3. I always enjoy visiting my family during vacations.
4. Revisiting my favorite destinations is always a lot of fun.

Question 10: Which of these music genres do you like best?
1. Country music reminds me of home.
2. Pop music always makes me want to dance.
3. I love the meaningful lyrics in Rhythm and Blues music.
4. Rock music makes me feel free and let go.

Question 11: Which of these subjects is your favourite?
1. I like the objective truths of mathematics.
2. Studying psychology helps me understand myself and others.
3. Learning about the origins of human life is fascinating.
4. I love learning about poetry and novels in my classes.

Question 12: How often do you post on social media?
1. I am far too busy to think about social media.
2. I only post on social media once in a great while.
3. I post on social media when interesting things happen in my life.
4. I use social media to share moments of my daily life with family and friends.

Question 13: Which of these historical figures inspires you the most?
1. As the symbol of freedom for everyone in the world, Mahatma Ghandi inspires me a lot.
2. If it weren't for Abraham Lincoln, we might be living in a very different world right now.
3. Martin Luther King Jr's I have a dream speech is a true inspiration for me.
4. As a science student aspiring to be a researcher, Albert Einstein is definitely my role model.

Question 14: How many siblings do you have?
1. I do not have any siblings.
2. I have one sibling.
3. I have two siblings.
4. I have more than two siblings.

Question 15: If you were deserted on an island (with basic living conditions met), which of these items would you most like to have?
1. I would definitely want my phone with me if I were on a deserted island.
2. If I were on a deserted island, my gaming device would help me pass the time.
3. Books would help me cope with being on a deserted island.
4. My pet would help me chase away lonely feelings on a deserted island.

Question 16: How regularly do you see your best friend?
1. I don't go a day without seeing my best friend.
2. I connect with my best friend at least once a week.
3. Last time I saw my best friend was about a month ago.
4. My best friend lives far away so I only get to see them once this year.

Question 17: Which of these new skills would you most like to attain?
1. I want to learn to speak French.
2. I want to learn to play the guitar.
3. I'd like to become a competent public speaker.
4. Skiing is a skill I'd love to pick up.

Question 18: Would you consider getting tattoos or piercings?
1. I have piercings and tattoos.
2. I would consider a discrete tattoo or piercing.
3. I never really thought tattoos or piercings were attractive.
4. I don't like needles so I would never get piercings or tattoos.

Question 19: Which of these would you most like to do if you won a million dollars?
1. If I won the lottery, I would use the money to see the world.
2. If I won the lottery, I would pay off my student debt.
3. A nice house and a new car are what a lottery win would buy me.
4. If I won the lottery, I would save the money for a rainy day.

Question 20: Which of these leisure activities is your favourite?
1. I enjoy watching Netflix in my free time.
2. I usually find myself working out at the gym in my free time.
3. I always look forward to going out with friends in my free time.
4. I would happily spend all my free time reading.

High Self-Disclosure Depth Questions:

Question 21: Under which circumstances would you consider cheating in an exam?
1. When it comes to exams, the chances of getting caught are so high that I would never cheat.
2. For people who do not prepare, cheating on an exam may be the best form of 'preparation'.
3. If a test determines whether I pass or fail a course, then I would cheat without a second thought.
4. If I do not believe a test will be fair then I don't feel the need to play by the rules.

Question 22: Which of these items do you fear most?
1. It creeps me out when bugs fly near my ear.
2. I think its terrifying the way snakes shed their skin.
3. When I get too high above the ground my stomach curls up.
4. The thought of going into deep water makes me feel sick.

Question 23: Which of these qualities would you most value in a friend?
1. The ability to keep a secret, no matter what, is the best quality in a friend.
2. A true friend is always someone who is there when you need them.
3. To me, a friend is someone who is always honest even when the truth might hurt me.
4. A friend is someone who understands my feelings and always shows me love and support.

Question 24: Which of these best describes your feelings about gossip?
1. I love hearing gossiping and being the one who tells it.
2. Gossip doesn't really move me one way or another.
3. I do not like talking about others behind their backs.
4. I sometimes like gossiping about others, but sometimes feel bad about it.

Question 25: Which of these words best describes your personality?
1. Life is more fun when you can laugh about it.
2. I like a quiet conversation with just one person.
3. I love socializing and meeting new people.
4. I really strive to achieve my goals.

Question 26: Which of these habits would you most want to break?
1. I am always procrastinating and would love to change that.
2. I should adopt a healthier sleeping schedule so that I feel sharper.
3. I need to reduce the junk food and start eating a healthier diet.

4. I really want to break the habit of biting my nails when I'm bored or under stress.

Question 27: With which of these political positions do you most identify?
1. I most agree with the social justice ideals of the liberal party.
2. I have a strong preference for the tradition and order that the conservative party discusses.
3. Politics? I avoid that, too many arguments.
4. I often vote for fringe parties when I have the choice to in elections.

Question 28: Which of these family members do you most admire?
1. The family member I admire the most is my mother.
2. In my family, I admire my dad the most.
3. I admire and feel close to all my family members.
4. I have always been closest to my grandparents.

Question 29: Which of these statements best describes your current relationship status?
1. I am single and loving it.
2. I enjoy dating, but not very seriously.
3. I have been in a solid relationship for the past few years.
4. My relationship status is a bit hard to explain.

Question 30: Which of these things is the most stressful to you?
1. I spend most of my time studying and it causes a lot of stress.
2. I think it's hard to plan for the future when there's so much I can't control.
3. It stresses me out when I argue with my friends.
4. I always feel the stress of not having enough money.

Question 31: With which of these religious beliefs do you most identify?
1. I go to church most Sundays.
2. I'm a non-Christian.
3. I don't think God exists.
4. I don't hold any specific religious beliefs.

Question 32: Which of these is the most important quality that you have or want to have as a person?
1. Honesty makes life easier, and more enjoyable.
2. Loyalty is one of the greatest qualities a person can possess.
3. The world would be a better place if people were kind above all else.
4. I think having an open mind is a great way to live life.

Question 33: Do you think you want a family with kids in the future?
1. Kids don't really fit with my lifestyle.

2. I might want kids someday, but I'm not too sure.
3. I really like kids but it will be a while before I feel ready.
4. I can't wait to start a family.

Question 34: When you are stressed, which of the following most helps you unwind?
1. A movie binge is the best way to end a hard day.
2. Music keeps me sane when times are tough.
3. Going for a run at the end of a long day is the best.
4. My friends always have my back when I need to talk about my day.

Question 35: Which of these best describes your greatest accomplishment so far?
1. It was one of the best moments in my life when I opened my admission letter from Western.
2. I won a prestigious scholarship and am pretty proud of it.
3. Living independently from my family is probably my biggest accomplishment so far.
4. I'm most proud of my ability to win athletic competitions.

Question 36: With which of these issues do you struggle most on a day-to-day basis?
1. It's very hard for me to stick to a schedule.
2. I always feel like I'm behind on my readings.
3. I find it difficult to talk to people I don't know very well.
4. I feel I'm a generally anxious person.

Question 37: Which of these things is most important to you in your life?
1. My family is the single most important thing in my life.
2. My future academic success is what most motivates me.
3. I think happiness is the thing that provides most meaning to life.
4. I feel that my beliefs are the basis of who I am.

Question 38: Which of these statements best describes your relationship with your parents?
1. I get along with my parents as if they were close friends.
2. My parents and I get along well but we don't really talk as friends.
3. I try to keep conversations with my parents on a pretty superficial level.
4. My parents and I can't seem to be in the same room for very long before we start to argue.

Question 39: How are you doing in school right now?
1. My academics this semester aren't going very well.
2. I'm doing OK in school - but I could be doing better.
3. My semester has been going pretty well, school-wise.

4. I think I'm doing super-well in my classes this term.

Question 40: Which of the following skills/characteristics would you most want to improve/change about yourself?
1. I wish I were more comfortable expressing my opinions to others.
2. I wish my body was a bit more athletic looking.
3. I wish I had better self-confidence.
4. If I had better self-control I would be really happy with things.

Appendix B: Experiments 1 to 5 Model Summaries

Table 7 Experiment 1 Type III Analysis of Deviance Table with Wald Chi-Square Tests (for cumulative link models)

		df	χ^2	*p*
The degree to which they are similar to you				
	Similarity	2	26.824	<.001
	Disclosure Depth	1	0.611	.434
	Number of Clicks	1	68.087	<.001
	Similarity * Disclosure Depth	2	0.393	.822
How well you think you've gotten to know them				
	Similarity	2	17.618	<.001
	Disclosure Depth	1	1.869	.172
	Number of Clicks	1	183.786	<.001
	Similarity * Disclosure Depth	2	0.518	.772
How comfortable you'd feel asking them for advice				
	Similarity	2	21.743	<.001
	Disclosure Depth	1	0.016	.899
	Number of Clicks	1	47.446	<.001
	Similarity * Disclosure Depth	2	2.560	.278
How much you would like to admit them to your circle of friends				
	Similarity	2	29.575	<.001
	Disclosure Depth	1	0.320	.572
	Number of Clicks	1	40.738	<.001
	Similarity * Disclosure Depth	2	0.843	.656
How much you'd like to actually meet them				
	Similarity	2	25.293	<.001
	Disclosure Depth	1	2.280	0131
	Number of Clicks	1	62.543	<.001
	Similarity * Disclosure Depth	2	0.634	.728

How trustworthy you think they are				
	Similarity	2	10.542	0.005
	Disclosure Depth	1	0.630	0.428
	Number of Clicks	1	57.497	<.001
	Similarity * Disclosure Depth	2	1.836	0.399
How friendly you think they are				
	Similarity	2	14.848	<.001
	Disclosure Depth	1	0.280	0.597
	Number of Clicks	1	49.362	<.000
	Similarity * Disclosure Depth	2	1.773	0.412

Table 8 Experiment 1 Cumulative-link Model Summary

		Beta	Std.Error	Z-Value	Odds Ratio	OR 95%CI Lower	OR 95%CI Upper
The degree to which they are similar to you							
	High Similarity	0.536	0.082	6.505***	1.708	1.455	2.009
	Low Similarity	-0.471	0.081	-5.804***	0.624	0.532	0.732
	High Disclosure Depth	-0.069	0.057	-1.207	0.933	0.835	1.044
	Number of Clicks	0.075	0.009	8.252***	1.077	1.059	1.097
	High Similarity * High Disclosure Depth	-0.010	0.081	-0.118	0.990	0.845	1.162
	Low Similarity * High Disclosure Depth	-0.038	0.080	-0.470	0.963	0.823	1.127
How well you think you've gotten to know them							
	High Similarity	0.428	0.083	5.168***	1.534	1.305	1.805
	Low Similarity	-0.269	0.081	-3.337***	0.764	0.652	0.895
	High Disclosure Depth	0.093	0.057	1.621	1.098	0.981	1.228
	Number of Clicks	0.128	0.009	13.557***	1.136	1.116	1.158
	High Similarity * High Disclosure Depth	0.046	0.082	0.558	1.047	0.891	1.229
	Low Similarity * High Disclosure Depth	-0.054	0.080	-0.674	0.947	0.809	1.109
How comfortable you'd feel asking them for advice							
	High Similarity	0.360	0.081	4.424***	1.433	1.222	1.682
	Low Similarity	-0.433	0.081	-5.323***	0.649	0.553	0.761
	High Disclosure Depth	-0.111	0.057	-1.959*	0.895	0.800	1.000
	Number of Clicks	0.061	0.009	6.888***	1.063	1.045	1.082
	High Similarity * High Disclosure Depth	0.124	0.081	1.535	1.132	0.966	1.326
	Low Similarity * High Disclosure Depth	-0.031	0.080	-0.388	0.969	0.828	1.134
How much you would like to admit them to your circle of friends							
	High Similarity	0.551	0.081	6.768***	1.735	1.480	2.036
	Low Similarity	-0.444	0.081	-5.470***	0.642	0.547	0.752
	High Disclosure Depth	-0.073	0.057	-1.287	0.929	0.831	1.039
	Number of Clicks	0.057	0.009	6.383***	1.059	1.040	1.077
	High Similarity * High Disclosure Depth	0.018	0.080	0.224	1.018	0.870	1.191
	Low Similarity * High Disclosure Depth	-0.071	0.080	-0.885	0.931	0.796	1.090

How much you’d like to actually meet them							
	High Similarity	0.610	0.083	7.357***	1.840	1.565	2.166
	Low Similarity	-0.499	0.081	-6.120***	0.607	0.517	0.712
	High Disclosure Depth	-0.087	0.057	-1.525	0.917	0.820	1.025
	Number of Clicks	0.071	0.009	7.908***	1.073	1.055	1.092
	High Similarity * High Disclosure Depth	-0.064	0.081	-0.793	0.938	0.799	1.099
	Low Similarity * High Disclosure Depth	0.037	0.080	0.461	1.038	0.886	1.215
How trustworthy you think they are							
	High Similarity	0.268	0.081	3.327***	1.307	1.117	1.531
	Low Similarity	-0.176	0.080	-2.198*	0.838	0.716	0.981
	High Disclosure Depth	-0.010	0.057	-0.172	0.990	0.886	1.107
	Number of Clicks	0.068	0.009	7.583***	1.070	1.052	1.089
	High Similarity * High Disclosure Depth	0.088	0.080	1.093	1.092	0.933	1.278
	Low Similarity * High Disclosure Depth	-0.099	0.080	-1.240	0.905	0.774	1.059
How friendly you think they are							
	High Similarity	0.369	0.082	4.523***	1.446	1.233	1.698
	Low Similarity	-0.290	0.080	-3.613***	0.748	0.639	0.876
	High Disclosure Depth	-0.030	0.057	-0.523	0.971	0.868	1.085
	Number of Clicks	0.062	0.009	7.026***	1.064	1.046	1.083
	High Similarity * High Disclosure Depth	0.083	0.081	1.020	1.086	0.927	1.273
	Low Similarity * High Disclosure Depth	0.018	0.080	0.226	1.018	0.871	1.191

Note. *: *p*<.05; **: *p*<.01; ***: *p*<.001

Table 9 Experiment 2 Type III Analysis of Deviance Table with Wald Chi-Square Tests (for Cumulative-Link Models)

		df	χ^2	*p*
The degree to which they are similar to you				
	Similarity	2	49.998	<.001
	Disclosure Depth	1	0.1501	.699
	Number of Clicks	1	111.767	<.001
	Similarity * Disclosure Depth	2	2.3084	.315
How well you think you've gotten to know them				
	Similarity	2	10.875	.004
	Disclosure Depth	1	0.179	.672
	Number of Clicks	1	148.725	<.001
	Similarity * Disclosure Depth	2	0.691	.707
How comfortable you'd feel asking them for advice				
	Similarity	2	23.143	<.001
	Disclosure Depth	1	0.090	0.764
	Number of Clicks	1	116.670	<.001
	Similarity * Disclosure Depth	2	0.442	0.802
How much you would like to admit them to your circle of friends				
	Similarity	2	33.090	<.001
	Disclosure Depth	1	1.386	.239
	Number of Clicks	1	103.402	<.001
	Similarity * Disclosure Depth	2	1.197	.550
How much you'd like to actually meet them				
	Similarity	2	21.705	<.001
	Disclosure Depth	1	0.107	.743
	Number of Clicks	1	132.579	<.001
	Similarity * Disclosure Depth	2	1.002	.606
How trustworthy you think they are				
	Similarity	2	23.033	<.001
	Disclosure Depth	1	2.262	.133
	Number of Clicks	1	68.935	<.001

	Similarity * Disclosure Depth	2	4.352	.114
How friendly you think they are				
	Similarity	2	15.084	<.001
	Disclosure Depth	1	0.257	.613
	Number of Clicks	1	75.621	<.001
	Similarity * Disclosure Depth	2	2.734	.255

Table 10 Experiment 2 Cumulative-Link Model Summary

		Beta	Std.Error	Z-Value	Odds Ratio	OR 95% CI Lower	OR 95% CI Upper
The degree to which they are similar to you							
	High Similarity	1.031	0.118	8.709***	2.804	2.227	3.544
	Low Similarity	-0.721	0.112	-6.392***	0.486	0.389	0.606
	High Disclosure Depth	0.006	0.077	0.077	1.006	0.865	1.169
	Number of Clicks	0.138	0.013	10.572***	1.148	1.119	1.178
	High Similarity * High Disclosure Depth	-0.059	0.110	-0.532	0.943	0.759	1.171
	Low Similarity * High Disclosure Depth	-0.103	0.108	-0.949	0.902	0.729	1.116
How well you think you’ve gotten to know them							
	High Similarity	0.601	0.114	5.244***	1.824	1.459	2.286
	Low Similarity	-0.373	0.109	-3.413***	0.689	0.555	0.853
	High Disclosure Depth	0.028	0.077	0.363	1.028	0.885	1.196
	Number of Clicks	0.162	0.013	12.195***	1.176	1.146	1.208
	High Similarity * High Disclosure Depth	-0.087	0.111	-0.779	0.917	0.737	1.141
	Low Similarity * High Disclosure Depth	0.016	0.107	0.150	1.016	0.823	1.255
How comfortable you’d feel asking them for advice							
	High Similarity	0.703	0.114	6.151***	2.019	1.616	2.530
	Low Similarity	-0.516	0.111	-4.653***	0.597	0.480	0.741
	High Disclosure Depth	-0.009	0.076	-0.114	0.991	0.854	1.151
	Number of Clicks	0.141	0.013	10.801	1.151	1.123	1.182
	High Similarity * High Disclosure Depth	0.050	0.110	0.453	1.051	0.847	1.305
	Low Similarity * High Disclosure Depth	0.019	0.108	0.177	1.019	0.825	1.260
How much you would like to admit them to your circle of friends							
	High Similarity	0.820	0.115	7.138***	2.270	1.815	2.847
	Low Similarity	-0.646	0.112	-5.771***	0.524	0.420	0.652
	High Disclosure Depth	0.077	0.076	1.007	1.080	0.923	1.255
	Number of Clicks	0.132	0.013	10.169***	1.142	1.113	1.171
	High Similarity * High Disclosure Depth	0.083	0.110	0.755	1.086	0.876	1.347
	Low Similarity * High Disclosure Depth	0.030	0.109	0.277	1.031	0.832	1.277

How much you’d like to actually meet them							
	High Similarity	0.772	0.117	6.622***	2.164	1.724	2.724
	Low Similarity	-0.568	0.111	-5.124***	0.567	0.456	0.704
	High Disclosure Depth	-0.013	0.077	-0.165	0.987	0.849	1.148
	Number of Clicks	0.153	0.013	11.514***	1.166	1.136	1.197
	High Similarity * High Disclosure Depth	-0.033	0.112	-0.296	0.968	0.777	1.204
	Low Similarity * High Disclosure Depth	0.105	0.108	0.970	1.110	0.899	1.372
How trustworthy you think they are							
	High Similarity	0.456	0.111	4.098***	1.578	1.270	1.964
	Low Similarity	-0.499	0.110	-4.534***	0.607	0.489	0.753
	High Disclosure Depth	-0.012	0.076	-0.154	0.989	0.852	1.146
	Number of Clicks	0.103	0.012	8.303***	1.108	1.082	1.136
	High Similarity * High Disclosure Depth	0.213	0.108	1.964*	1.237	1.001	1.531
	Low Similarity * High Disclosure Depth	-0.173	0.107	-1.613	0.841	0.682	1.038
How friendly you think they are							
	High Similarity	0.564	0.112	5.058***	1.758	1.414	2.190
	Low Similarity	-0.491	0.111	-4.437***	0.612	0.493	0.760
	High Disclosure Depth	0.024	0.076	0.314	1.024	0.883	1.188
	Number of Clicks	0.108	0.012	8.696***	1.114	1.087	1.142
	High Similarity * High Disclosure Depth	0.044	0.109	0.407	1.045	0.844	1.294
	Low Similarity * High Disclosure Depth	0.125	0.108	1.153	1.133	0.916	1.402

Note. *: *p*<.05; **: *p*<.01; ***: *p*<.001

Table 11 Experiment 3 Type III Analysis of Deviance Table with Wald Chi-Square Tests (for Cumulative Link Models)

		df	χ^2	*p*
The degree to which they are similar to you				
	Similarity	2	19.815	<.001
	Disclosure Depth	1	0.254	.614
	Number of Clicks	1	56.727	<001
	Similarity * Disclosure Depth	2	0.687	.709
How well you think you've gotten to know them				
	Similarity	2	7.222	.027
	Disclosure Depth	1	0.006	.937
	Number of Clicks	1	83.515	<.001
	Similarity * Disclosure Depth	2	1.187	.552
How comfortable you'd feel asking them for advice				
	Similarity	2	18.389	<.001
	Disclosure Depth	1	1.519	.218
	Number of Clicks	1	57.294	<.001
	Similarity * Disclosure Depth	2	1.268	.531
How much you would like to admit them to your circle of friends				
	Similarity	2	22.892	<.001
	Disclosure Depth	1	2.174	.140
	Number of Clicks	1	63.756	<.001
	Similarity * Disclosure Depth	2	2.964	.227
How much you'd like to actually meet them				
	Similarity	2	17.285	<.001
	Disclosure Depth	1	0.751	.386
	Number of Clicks	1	72.473	<.001
	Similarity * Disclosure Depth	2	1.029	.598
How trustworthy you think they are				
	Similarity	2	5.926	.052
	Disclosure Depth	1	0.333	.564
	Number of Clicks	1	52.778	<.001

	Similarity * Disclosure Depth	2	1.748	.417
How friendly you think they are				
	Similarity	2	7.621	.022
	Disclosure Depth	1	1.532	.216
	Number of Clicks	1	39.097	<.001
	Similarity * Disclosure Depth	2	4.742	.093
How much you think they'd like to be your friend				
	Similarity	2	27.336	<.001
	Disclosure Depth	1	3.746	.053
	Number of Clicks	1	55.154	<.001
	Similarity * Disclosure Depth	2	2.889	.237

Table 12 Experiment 3 Cumulative-Link Model Summary

		Beta	Std.Error	Z-Value	Odds Ratio	OR 95% CI Lower	OR 95% CI Upper
The degree to which they are similar to you							
	High Similarity	0.819	0.120	6.854***	2.269	1.798	2.873
	Low Similarity	-0.602	0.117	-5.170***	0.548	0.435	0.687
	High Disclosure Depth	-0.002	0.080	-0.025	0.998	0.853	1.167
	Number of Clicks	0.102	0.013	7.532***	1.107	1.078	1.137
	High Similarity * High Disclosure Depth	0.073	0.114	0.638	1.076	0.860	1.346
	Low Similarity * High Disclosure Depth	0.015	0.113	0.134	1.015	0.814	1.267
How well you think you’ve gotten to know them							
	High Similarity	0.542	0.119	4.562***	1.720	1.364	2.174
	Low Similarity	-0.394	0.114	-3.445***	0.675	0.539	0.843
	High Disclosure Depth	0.100	0.080	1.244	1.105	0.944	1.294
	Number of Clicks	0.127	0.014	9.139***	1.136	1.105	1.167
	High Similarity * High Disclosure Depth	-0.111	0.115	-0.964	0.895	0.713	1.122
	Low Similarity * High Disclosure Depth	0.104	0.112	0.925	1.109	0.891	1.382
How comfortable you’d feel asking them for advice							
	High Similarity	0.569	0.117	4.880***	1.766	1.407	2.221
	Low Similarity	-0.451	0.115	-3.932***	0.637	0.508	0.797
	High Disclosure Depth	0.050	0.080	0.625	1.051	0.899	1.230
	Number of Clicks	0.102	0.014	7.569***	1.108	1.079	1.138
	High Similarity * High Disclosure Depth	0.122	0.114	1.074	1.130	0.904	1.413
	Low Similarity * High Disclosure Depth	-0.026	0.112	-0.233	0.974	0.782	1.213
How much you would like to admit them to your circle of friends							
	High Similarity	0.606	0.117	5.171***	1.832	1.458	2.308
	Low Similarity	-0.426	0.115	-3.714***	0.653	0.521	0.817
	High Disclosure Depth	0.042	0.080	0.527	1.043	0.892	1.221
	Number of Clicks	0.109	0.014	7.985***	1.115	1.086	1.146
	High Similarity * High Disclosure Depth	0.165	0.115	1.443	1.180	0.943	1.477
	Low Similarity * High Disclosure Depth	0.012	0.113	0.102	1.012	0.811	1.261

How much you’d like to actually meet them							
	High Similarity	0.613	0.118	5.206***	1.845	1.467	2.327
	Low Similarity	-0.399	0.115	-3.466***	0.670	0.535	0.840
	High Disclosure Depth	0.058	0.080	0.718	1.059	0.905	1.240
	Number of Clicks	0.117	0.014	8.513***	1.124	1.095	1.155
	High Similarity * High Disclosure Depth	0.065	0.115	0.566	1.067	0.852	1.336
	Low Similarity * High Disclosure Depth	0.050	0.113	0.444	1.051	0.843	1.312
How trustworthy you think they are							
	High Similarity	0.260	0.113	2.308*	1.297	1.040	1.618
	Low Similarity	-0.337	0.115	-2.931**	0.714	0.569	0.894
	High Disclosure Depth	-0.032	0.080	-0.394	0.969	0.828	1.133
	Number of Clicks	0.097	0.013	7.265***	1.102	1.074	1.132
	High Similarity * High Disclosure Depth	0.109	0.112	0.976	1.115	0.896	1.389
	Low Similarity * High Disclosure Depth	0.035	0.113	0.312	1.036	0.830	1.294
How friendly you think they are							
	High Similarity	0.329	0.115	2.866**	1.389	1.110	1.741
	Low Similarity	-0.391	0.115	-3.403***	0.677	0.540	0.847
	High Disclosure Depth	0.051	0.080	0.641	1.052	0.900	1.231
	Number of Clicks	0.082	0.013	6.253***	1.085	1.058	1.114
	High Similarity * High Disclosure Depth	0.120	0.113	1.062	1.128	0.904	1.408
	Low Similarity * High Disclosure Depth	0.127	0.113	1.123	1.135	0.910	1.416
How much you think they’d like to be your friend							
	High Similarity	0.599	0.118	5.091***	1.820	1.446	2.294
	Low Similarity	-0.511	0.117	-4.987***	0.600	0.477	0.753
	High Disclosure Depth	0.083	0.081	1.036	1.086	0.929	1.271
	Number of Clicks	0.100	0.013	7.427***	1.105	1.077	1.135
	High Similarity * High Disclosure Depth	0.187	0.114	1.647	1.206	0.965	1.508
	Low Similarity * High Disclosure Depth	-0.134	0.113	-1.183	0.875	0.700	1.092

Note. *: *p*<.05; **: *p*<.01; ***: *p*<.001

Table 13 Experiment 4 Type III Analysis of Variance table with Satterthwaite's method (for Linear-Mixed Models)

		Numerator Df	Denominator DF	*F*-value	*p*
Social Preferences (aggregated means)					
	Similarity	2	701.728	0.205	.815
	Disclosure Depth	1	699.676	0.040	.841
	Personalistic Attribution	1	139.872	0.000	.994
	Number of Clicks	1	793.173	0.322	.570
	Similarity * Disclosure Depth	2	699.910	1.174	.310
	Personalistic * Disclosure Depth	1	699.756	0.015	.902
Perceived Desirable Traits (aggregated means)					
	Similarity	2	701.163	3.875	.021
	Disclosure Depth	1	699.732	0.050	.823
	Personalistic Attribution	1	139.868	0.729	.395
	Number of Clicks	1	767.054	0.632	.427
	Similarity * Disclosure Depth	2	699.895	0.547	.579
	Personalistic * Disclosure Depth	1	699.787	0.288	.592

Table 14 Experiment 4 Linear-Mixed Model Summary

		Beta	Std. Error	*df*	*t*-value	*p*
Social Preferences (aggregated means)						
	(Intercept)	4.246	0.085	611.251	49.780	<.001
	High Similarity	-0.026	0.048	701.698	-0.545	.586
	Low Similarity	0.000	0.048	703.543	-0.008	.994
	High Disclosure Depth	0.007	0.033	699.676	0.201	.841
	High Personalistic Attribution	0.000	0.050	139.872	-0.008	.994
	Number of Clicks	0.003	0.006	793.173	0.568	.570
	High Similarity * High Disclosure Depth	0.015	0.047	700.109	0.322	.748
	Low Similarity * High Disclosure Depth	0.054	0.047	699.702	1.136	.256
	High Personalistic * High Disclosure Depth	-0.004	0.033	699.756	-0.123	.902
Perceived Desirable Traits (aggregated means)						
	(Intercept)	4.605	0.068	511.590	67.755	<.001
	High Similarity	-0.073	0.035	701.143	-2.100	.036
	Low Similarity	-0.017	0.035	702.431	-0.492	.623
	High Disclosure Depth	-0.005	0.024	699.732	-0.223	.823
	High Personalistic Attribution	-0.038	0.045	139.868	-0.854	.395
	Number of Clicks	0.003	0.004	767.054	0.795	.427
	High Similarity * High Disclosure Depth	0.005	0.034	700.034	0.155	.877
	Low Similarity * High Disclosure Depth	0.028	0.034	699.750	0.818	.414
	High Personalistic * High Disclosure Depth	-0.013	0.024	699.787	-0.536	.592

Table 15 . Experiment 4 Type III Analysis of Deviance Table with Wald Chi-Square Tests (for Cumulative Link Models)

		df	χ^2	*p*
The degree to which they are similar to you				
	Similarity	2	0.687	.709
	Disclosure Depth	1	0.909	.340
	Personalistic Attribution	1	0.072	.788
	Number of Clicks	1	0.026	.872
	Similarity * Disclosure Depth	2	1.302	.522
	Personalistic * Disclosure Depth	1	1.055	.304
How well you think you've gotten to know them				
	Similarity	2	1.572	.456
	Disclosure Depth	1	0.166	.684
	Personalistic Attribution	1	3.132	.077
	Number of Clicks	1	1.238	.266
	Similarity * Disclosure Depth	2	3.120	.210
	Personalistic * Disclosure Depth	1	0.035	.853
How much you'd like to actually meet them				
	Similarity	2	45.474	<.001
	Disclosure Depth	1	2.983	.084
	Personalistic Attribution	1	0.111	.739
	Number of Clicks	1	57.459	<.000
	Similarity * Disclosure Depth	2	5.145	.076
	Personalistic * Disclosure Depth	1	0.379	.538

Table 16 Experiment 4 Cumulative-Link Model Summary

		Beta	Std. Error	Z-Value	Odds Ratio	OR 95% CI Lower	OR 95% CI Upper
The degree to which they are similar to you							
	High Similarity	-0.027	0.087	-0.317	0.973	0.821	1.153
	Low Similarity	-0.041	0.088	-0.472	0.959	0.808	1.140
	High Disclosure Depth	-0.058	0.061	-0.954	0.944	0.837	1.063
	High Personalistic Attribution	-0.017	0.061	-0.279	0.983	0.873	1.108
	Number of Clicks	0.003	0.010	0.250	1.003	0.983	1.023
	High Similarity * High Disclosure Depth	0.066	0.086	0.761	1.068	0.902	1.265
	Low Similarity * High Disclosure Depth	0.030	0.087	0.344	1.030	0.869	1.221
	High Personalistic * High Disclosure Depth	-0.063	0.061	-1.027	0.939	0.833	1.059
How well you think you’ve gotten to know them							
	High Similarity	-0.085	0.089	-0.959	0.918	0.771	1.093
	Low Similarity	-0.018	0.089	-0.203	0.982	0.824	1.170
	High Disclosure Depth	-0.026	0.062	-0.411	0.975	0.863	1.101
	High Personalistic Attribution	-0.113	0.063	-1.814	0.893	0.790	1.009
	Number of Clicks	0.012	0.011	1.179	1.013	0.992	1.034
	High Similarity * High Disclosure Depth	0.050	0.089	0.565	1.051	0.884	1.251
	Low Similarity * High Disclosure Depth	0.103	0.088	1.167	1.108	0.933	1.317
	High Personalistic * High Disclosure Depth	-0.012	0.062	-0.186	0.988	0.875	1.117
How much you’d like to actually meet them							
	High Similarity	0.627	0.121	5.185***	1.871	1.478	2.375
	Low Similarity	-0.434	0.116	-3.725***	0.648	0.515	0.814
	High Disclosure Depth	0.137	0.082	1.675	1.147	0.977	1.348
	High Personalistic Attribution	-0.041	0.082	-0.503	0.960	0.817	1.127
	Number of Clicks	0.114	0.015	7.454***	1.120	1.088	1.155
	High Similarity * High Disclosure Depth	-0.197	0.117	-1.681	0.821	0.652	1.033
	Low Similarity * High Disclosure Depth	0.248	0.115	2.163*	1.282	1.024	1.607
	High Personalistic * High Disclosure Depth	0.051	0.082	0.615	1.052	0.896	1.236

Table 17 Experiment 5 Type III Analysis of Variance Table with Satterthwaite's Method (for Linear-Mixed Models)

		Numerator Df	Denominator DF	*F*-value	*p*
Perceived Reciprocity (aggregated means)					
	Similarity	1	520	6.733	.010
	Reciprocity	2	520	7.055	.001
	Similarity * Reciprocity	2	520	0.136	.873
Social Preferences (aggregated means)					
	Similarity	1	520	30.269	<.001
	Reciprocity	2	520	2.017	.134
	Similarity * Reciprocity	2	520	0.640	.528
Perceived Desirable Traits (aggregated means)					
	Similarity	1	520	3.093	.079
	Reciprocity	2	520	2.519	.082
	Similarity * Reciprocity	2	520	0.403	.669
Participants' Number of Reciprocated Disclosure					
	Similarity	1	520	0.161	.689
	Reciprocity	2	520	2.927	.054
	Similarity * Reciprocity	2	520	0.306	.737

Table 18 Experiment 5 Linear-Mixed Model Summary

		Beta	Std. Error	*df*	*t*-value	*p*
Perceived Reciprocity (aggregated means)						
	(Intercept)	5.007	0.071	104	70.873	<.001
	High Similarity	0.103	0.040	520	2.595	.010
	High Reciprocity	0.212	0.056	520	3.755	<.001
	Low Reciprocity	-0.101	0.056	520	-1.792	.074
	High Similarity* High Reciprocity	0.010	0.056	520	0.171	.865
	High Similarity* Low Reciprocity	-0.029	0.056	520	-0.512	.609
Social Preferences (aggregated means)						
	(Intercept)	4.846	0.071	104	68.015	<.001
	High Similarity	0.171	0.031	520	5.502	<.001
	High Reciprocity	0.088	0.044	520	1.991	.047
	Low Reciprocity	-0.054	0.044	520	-1.227	.220
	High Similarity* High Reciprocity	-0.048	0.044	520	-1.082	.280
	High Similarity* Low Reciprocity	0.036	0.044	520	0.827	.409
Perceived Desirable Traits (aggregated means)						
	(Intercept)	4.833	0.066	104	73.084	<.001
	High Similarity	0.038	0.022	520	1.759	.079
	High Reciprocity	0.061	0.031	520	1.969	.049
	Low Reciprocity	-0.059	0.031	520	-1.917	.056
	High Similarity* High Reciprocity	-0.024	0.031	520	-0.777	.437
	High Similarity* Low Reciprocity	0.024	0.031	520	0.777	.437
Participants’ Number of Reciprocated Disclosure						
	(Intercept)	4.394	0.059	104	75.015	<.001
	High Similarity	0.022	0.056	520	0.401	.689
	High Reciprocity	0.005	0.079	520	0.061	.952
	Low Reciprocity	-0.168	0.079	520	-2.125	.034
	High Similarity* High Reciprocity	0.021	0.079	520	0.263	.793
	High Similarity* Low Reciprocity	0.040	0.079	520	0.506	.613

Table 19 Experiment 5 Type III Analysis of Deviance Table with Wald Chi-Square Tests (for Cumulative Link Models)

		df	χ^2	*p*
The degree to which they are similar to you				
	Similarity	1	37.561	<.001
	Reciprocity	2	4.808	.090
	Similarity* Reciprocity	2	0.612	.736
How much you'd like to actually meet them				
	Similarity	1	0.018	.893
	Reciprocity	2	0.279	.870
	Similarity* Reciprocity	2	1.215	.545

Table 20 Experiment 5 Cumulative-Link Model Summary

		Beta	Std. Error	Z-Value	Odds Ratio	OR 95% CI Lower	OR 95% CI Upper
The degree to which they are similar to you							
	High Similarity	0.455	0.074	6.129***	1.576	1.364	1.824
	High Reciprocity	0.224	0.103	2.182*	1.251	1.023	1.531
	Low Reciprocity	-0.129	0.102	-1.272	0.879	0.720	1.073
	High Similarity * High Reciprocity	0.011	0.102	0.107	1.011	0.827	1.236
	High Similarity * Low reciprocity	0.063	0.102	0.621	1.065	0.873	1.300
How much you'd like to actually meet them							
	High Similarity	0.009	0.070	0.135	1.010	0.879	1.159
	High Reciprocity	-0.049	0.099	-0.498	0.952	0.784	1.156
	Low Reciprocity	0.009	0.099	0.091	1.009	0.831	1.226
	High Similarity * High Reciprocity	0.108	0.099	1.088	1.114	0.917	1.353
	High Similarity * Low reciprocity	-0.068	0.099	-0.686	0.934	0.769	1.135

Note. *: *p*<.05; **: *p*<.01; ***: *p*<.001

Appendix C: Experiment 7 ANOVA Results and Pair-Wise Comparisons

Table 21 Experiment 7 ANOVA Results on All Measures

		Effects
Number of Julie's Responses across Conditions (Experimenter Bias Check)[1]	Similarity:	*F*(2,138)=2.311, *p*=.103, η^2=.032
	Partner Identity:	*F*(1,138)=2.192, *p*=.141, η^2=.016
	Interaction:	*F*(2,138)=0.030, *p*=.970, η^2<.001
Similarity Manipulation Check	**Similarity:**	***F*(2,149)=8.614, *p*<.001, η^2=.104**
	Partner Identity:	*F*(1,149)=2.288, *p*=.132, η^2=.015
	Interaction:	*F*(2,149)=1.159, *p*=.317, η^2=.015
Perceived Warmth	Similarity:	*F*(2,149)=0.901, *p*=.408, η^2=.012
	Partner Identity:	*F*(1,149)=0.201, *p*=.655, η^2=.001
	Interaction:	***F*(2,149)=4.387, *p*=.014, η^2=.056**
Perceived Competence	Similarity:	*F*(2,149)=0.595, *p*=.553, η^2=.008
	Partner Identity:	*F*(1,149)=0.290, *p*=.591, η^2=.002
	Interaction:	*F*(2,149)=2.323, *p*=.102, η^2=.030
Mind Perception- Agency	Similarity:	*F*(2,149)=1.219, *p*=.298, η^2=.016
	Partner Identity:	*F*(1,149)=2.440, *p*=.120, η^2=.016
	Interaction:	***F*(2,149)=4.138, *p*=.018, η^2=.053**
Mind Perception- Experience	Similarity:	*F*(2,149)=0.100, *p*=.905, η^2=.001
	Partner Identity:	*F*(1,149)=3.606, *p*=.060, η^2=.024
	Interaction:	***F*(2,149)=3.668, *p*=.028, η^2=.047**
Anthropomorphism	Similarity:	*F*(2,149)=0.042, *p*=.958, η^2=.001
	Partner Identity:	*F*(1,149)<0.001, *p*=.990, η^2<.001
	Interaction:	*F*(2,149)=2.831, *p*=.062, η^2=.037
Liking and Rapport	Similarity:	*F*(2,149)=0.003, *p*=.997, η^2<.001
	Partner Identity:	*F*(1,149)=0.020, *p*=.889, η^2<.001
	Interaction:	*F*(2,149)=0.955, *p*=.387, η^2=.013
Conversation Length (in seconds)	Similarity:	*F*(2,148)=0.537, *p*=.586, η^2=.007
	Partner Identity:	*F*(1,148)=3.378, *p*=.068, η^2=.022
	Interaction:	*F*(2,148)=2.062, *p*=.131, η^2=.027
AU4 Activity	Similarity:	*F*(2,148)=1.335, *p*=.226, η^2=.018
	Partner Identity:	***F*(1,148)=8.815, *p*=.003, η^2=.056**
	Interaction:	*F*(2,148)=1.003, *p*=.369, η^2=.013
AU6 Activity	Similarity:	*F*(2,148)=0.228, *p*=.797, η^2=.003
	Partner Identity:	*F*(1,148)=0.005, *p*=.942, η^2<.001
	Interaction:	*F*(2,148)=1.498, *p*=.227, η^2=.020
AU12 Activity	Similarity:	*F*(2,148)=0.533, *p*=.588, η^2=.007
	Partner Identity:	*F*(1,148)=0.673, *p*=.413, η^2=.005
	Interaction:	*F*(2,148)=0.739, *p*=.479, η^2=.010

Note. Effects statistically significant at the .05 level are in bold. [1]*: Log files from 11 participants were lost due to technical failure.*

Table 22 Experiment 7 Mean and Standard Deviation by Cell and Pair-Wise Comparisons for Dependent Variables

		Low Similarity	**Medium Similarity**	**High Similarity**
Warmth	Human Partner	4.84 (*1.16*)	5.54 (*0.85*)	4.58 (*1.23*)
	AI Partner	5.11 (*1.01*)	4.64 (*1.46*)	4.96 (*1.35*)
	p-Value	.425	**.009**	.260
Competence	Human Partner	4.25 (*1.18*)	4.96 (*0.90*)	4.25 (*1.18)*
	AI Partner	4.39 (*1.30*)	4.20 (*1.57*)	4.49 (*1.42*)
	p-Value	.700	**.044**	.445
Mind Perception - Agency	Human Partner	4.13 (*1.44*)	4.78 (*0.90*)	3.91 (*1.07*)
	AI Partner	4.35 (*1.28*)	3.64 (*1.43*)	3.87 (*1.35*)
	p-Value	.525	**.002**	.922
Mind Perception - Experience	Human Partner	2.63 (*1.07*)	3.14 (*1.05*)	2.86 (*1.15*)
	AI Partner	2.89 (*1.16*)	2.21 (*1.26*)	2.50 (*0.92*)
	p-Value	.390	**.004**	.253
Anthropomorphism	Human Partner	3.00 (*1.06*)	3.48 (*1.31*)	3.00 (*1.01*)
	AI Partner	3.32 (*1.29*)	2.77 (*1.32*)	3.39 (*1.56*)
	p-Value	.364	.056	.294
Liking and Rapport	Human Partner	3.78 (*1.15*)	4.07 (*1.15*)	3.78 (*1.37*)
	AI Partner	4.02 (*1.34*)	3.69 (*1.34*)	4.00 (*1.46*)
	p-Value	.475	.302	.551
Conversation Length (in minutes)	Human Partner	15.09 (*3.28*)	15.50 (*2.87*)	14.85 (*2.75*)
	AI Partner	14.90 (*3.51*)	13.33 (*1.33*)	14.68 (*2.76*)
	p-Value	.811	**.008**	.830
AU4 Activity	Human Partner	0.02 (*0.04*)	0.01(*0.01*)	0.01(*0.03*)
	AI Partner	0.08 (0.19)	0.03 (*0.07*)	0.11(0.22)
	p-Value	.067	.518	**.008**
AU6 Activity	Human Partner	0.23(*0.29*)	0.28(*0.44*)	0.14(*0.14*)
	AI Partner	0.23(*0.35*)	0.18(*0.23*)	0.25(*0.40*)
	p-Value	.971	.244	.203
AU12 Activity	Human Partner	0.43(*0.42*)	0.53(*0.49*)	0.34(*0.40*)
	AI Partner	0.42(*0.56*)	0.34(*0.44*)	0.35(*0.42*)
	p-Value	.961	.150	.922

Note. Means and Standard Deviations (in brackets) for each cell across the dependent measures. The p-values are the significant levels for the pair-wise comparisons of means between the Human Partner and AI partner conditions at each level of Similarity, adjusted by Bonferroni method. Statistically significant p-values at the .05 level are in bold font.

Curriculum Vitae

Yixian Li

Department of Psychology,
the University of Western Ontario,
London, ON, Canada, N6A 5C2

Academic Information

Doctor of Philosophy, Psychology 2020
The University of Western Ontario

- Supervisor: Dr. Erin Heerey

Master of Science, Psychology 2014-2016
The University of Western Ontario

- Master's Thesis: Effects of Value Reasoning on Stigmatization of People with Schizophrenia
- Supervisor: Dr. Ross Norman

Bachelor of Art, Honor Specialization in Psychology 2009-2013
The University of Western Ontario

- Honors Thesis: Effects of Uncertainty Orientation and Social Intelligence on Self-Ingroup Association.
- Supervisor: Dr. Richard Sorrentino

Honors and Awards

Graduate Research Award	2019
Mitacs Globalink Research Award	2018
Ontario Graduate Scholarships	2018, 2019
Graduate Student Teaching Assistant Awards of Excellence Nomination	2017, 2018
Western Graduate Research Grant	2014-Present
Dean's Honor List	2009-2013

Publications

Li, Y. & Heerey, E. (2020). *Does Self-Disclosure Depth Really Matter in Developing Initial Feelings of Liking?* Manuscript submitted for publication.

Li, Y., Heerey, E., & Gratch, J. (2020). *Negative Perceptions of a Self-Disclosing AI: the Potential Role of the Uncanny Valley Effect.* Manuscript in preparation.

Li, Y., Babcock, S., Stewart, S. L., Hirdes, J. P., & Schwean, V. L. (2020). *Psychometric Evaluation of the Depressive Severity Index (DSI) among Children and Youth using the interRAI Child and Youth Mental Health (ChYMH) Assessment Tool.* Manuscript submitted for publication.

Stewart, S. L., Babcock, S. E., **Li, Y.**, & Dave, H. P. (2020). A psychometric evaluation of the interRAI Child and Youth Mental Health instruments (ChYMH) anxiety scale in children with and without developmental disabilities. *BMC psychiatry, 20*(1), 1-14.

Li, Y., Sorrentino, R. M., Norman, R. M., Hampson, E., & Ye, Y. (2017). Effects of symptom versus recovery video, similarity, and uncertainty orientation on the stigmatization of schizophrenia. *Personality and Individual Differences*, *106*, 117-121.

Babcock, S., **Li, Y.**, Sinclair, V. M., Thomson, C., & Campbell, L. (2017). Two replications of an investigation on empathy and utilitarian judgement across socioeconomic status. *Scientific data, 4*(1)*,* 1-11.

Norman, R. M., **Li, Y.**, Sorrentino, R., Hampson, E., & Ye, Y. (2017). The differential effects of a focus on symptoms versus recovery in reducing stigma of schizophrenia. *Social psychiatry and psychiatric epidemiology, 52*(11), 1385-1394.

Oral Presentations

Yixian Li. (May 15, 2019), *How do People React to Self-Disclosed Similarity from a Human or an AI?* Data Blitz talk at Waterloo-Western-Wilfrid Laurier Social Psychology 2019 Annual Conference, Waterloo, Ontario.

Yixian Li. (May 31, 2018), *Self-Disclosure and Interpersonal Liking*. Oral presentation at Waterloo-Western-Wilfrid Laurier Social Psychology 2018 Annual Conference, Waterloo, Ontario.

Yixian Li, & Ross Norman, (June 23, 2016). *Effect of Value Reasoning on Stigmatization of Schizophrenia.* Oral presentation at Academic Research Day, Psychiatry Department, the University of Western Ontario, London, Ontario.

Yixian Li, Ross Norman, Richard Sorrentino, Elizabeth Hampson, & Yang Ye. (May 14, 2014). *Can Informational Videos Reduce Stigma towards Schizophrenia?* Oral presentation at University of Western Ontario-University of Waterloo Social Psychology 2014 Annual Conference, London, Ontario.

Poster Presentations

Yixian Li, & Erin Heerey, (Feb. 7, 2019), *The Ingredients of Self-Disclosure: Similarity, Depth, Reciprocity, and Liking.* Poster presentation at Society for Personality and Social Psychology 20th Annual Convention, Portland, Oregon.

Yixian Li, & Erin Heerey, (May 25, 2018), *Good to Know You: Effects of Similarity and Disclosure Depth on Interpersonal Liking.* Poster presentation at Association for Psychological Science 30th Annual Convention, San Francisco, California.

Yixian Li, & Erin Heerey, (May 27, 2017), *Getting to Know You: Effects of Perceived Similarity and Self-Disclosure on the Development of Liking.* Poster presentation at Association for Psychological Science 29th Annual Convention, Boston, Massachusetts.

Yixian Li, & Ross Norman, (Jan. 30, 2016). *Arguing Against Self-Enhancement Values Leads to Less Mental Illness Stigma than Arguing for Self-Transcendence Values.* Poster Presentation at Society for Personality and Social Psychology 17th Annual Convention, San Diego, California.

Yixian Li, Ross Norman, Richard Sorrentino, Elizabeth Hampson, & Yang Ye, (Feb. 27, 2015). *Can Informational Videos Reduce Stigmas of Schizophrenia? The Content Matters.* Poster presentation at Society for Personality and Social Psychology 16th Annual Convention, Long Beach, California.

Yixian Li, Richard Sorrentino, & Yang Ye, (Feb. 15, 2014). *Effects of Uncertainty Orientation and Social Intelligence on Self-Ingroup Association.* Poster presentation at Society for Personality and Social Psychology 15th Annual Convention, Austin, Texas.

Guest Lectures and Invited Talks

Yixian Li. (Oct. 22, 2019), *Group Processes.* Guest lecture for Introduction to Social Psychology (Psychology 2720), the University of Western Ontario, London, Ontario.

Yixian Li. (Oct. 18, 2017), *Factorial Designs in ANOVA.* Guest lecture for Psychological Statistics Using Computers (Psychology 3800), the University of Western Ontario, London, Ontario.

Yixian Li. (March 5, 2014). *Prejudice and Discrimination.* Invited in-class talk for Introduction to Psychology (Psychology 1000), the University of Western Ontario, London, Ontario.