Published in final edited form as:

Comput Human Behav. 2016 February 1; 55 Pt B: 804–816. doi:10.1016/j.chb.2015.07.067.

Watch what I do, not what I say I do: Computer-based avatars to assess behavioral inhibition, a vulnerability factor for anxiety disorders

Catherine E. Myers[1,2,3], **John A. Kostek**[1], **Barbara Ekeh**[4], **Rosanna Sanchez**[3,4], **Yasheca Ebanks-Williams**[1], **Ann L. Krusznis**, **Noah Weinflash**[1], and **Richard J. Servatius**[2,5]

[1]Department of Veterans Affairs, VA New Jersey Health Care System, East Orange, NJ

[2]Department of Pharmacology, Physiology & Neuroscience, New Jersey Medical School, Rutgers The State University of New Jersey, Newark, NJ

[3]Department of Psychology, Rutgers The State University of New Jersey, Newark, NJ

[4]Honors College, Rutgers The State University of New Jersey, Newark, NJ

[5]Department of Veterans Affairs, Syracuse, NY

Abstract

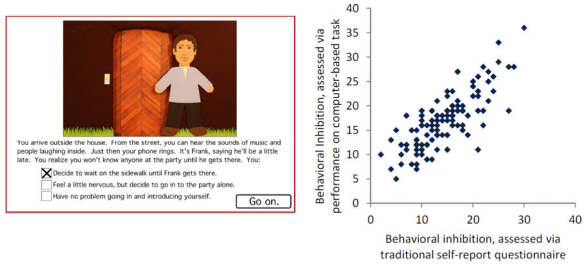
Behavioral inhibition (BI), a tendency to withdraw from or avoid novel social and non-social situations, is a personality trait which can confer risk for anxiety disorders. Like many personality traits, BI is often assessed via self-report questionnaires where respondents rate themselves for frequency of certain behaviors or feelings. However, questionnaires have inherent limitations, particularly in psychiatric populations where there may be unawareness of deficit. A viable alternative may be virtual environments, in which the participant guides an on-screen "avatar" through a series of onscreen events meant to simulate real-world situations. Here, we report on initial development of such an assessment tool, involving several onscreen scenarios with choice points where the participant can select from response options corresponding to inhibited or uninhibited behaviors. In two experiments involving over 300 college students, scores on the computer-based task were strongly correlated with BI scores attained through self-report questionnaire ($r>.780$, $p<.001$); this relationship held regardless of participant gender and experience with computer games. The results suggest that virtual environments may hold promise as alternative formats for assessment of personality traits in populations unsuited to traditional

Corresponding Author: Catherine E. Myers, PhD, Neurobehavioral Research Lab (NBRL), VA New Jersey Health Care System, 385 Tremont Ave., Mailstop 15, East Orange, NJ 07018 USA, Phone: (+1) 973-676-1000 x(1)1810, Catherine.Myers2@va.gov.

Contribution of authors: CEM and RJS devised the experimental design and hypothesis; CEM, RJS, JAK, BE, and ALK developed the specific task scenarios and designed the avatars and visual stimuli to accompany the script; JAK and CEM prepared the script, including task questions and possible responses; ALK constructed the avatars and the visual stimuli; CEM and NW programmed the software; BE and RS collected the behavioral and questionnaire data; CEM, YEW, BE and RS scored and analyzed the data; all authors (CEM, JAK, BE, RS, YEW, ALK, NW, RJS) contributed to interpretation and manuscript preparation.

paper-and-pencil questionnaire formats due to psychopathology, limited attention span, or poor vocabulary and/or literacy skills.

Graphical Abstract



Keywords

computer-based assessment; personality assessment tools; virtual environments; anxiety; behavioral inhibition

1. Introduction

The widespread use and development of online and virtual environments opens up new potential for a range of applications, such as studying individual participants' behavior in simulated situations. Prior work has examined how individuals create and express online personalities, and how such online personalities can change in different online settings (Guitton, 2010; Vasalou & Joinson, 2009). Other studies have considered how people use online identities that may be similar to, or different from, their real selves (Joinson & Dietz-Uhler, 2002).

The current study examines whether simple, highly-controlled online environments can be used to evoke behavior that is indicative of specific personality traits which have been implicated in vulnerability to psychiatric disorder. We focus on the personality trait of behavioral inhibition (BI), which confers risk for several psychiatric disorders, and which has traditionally been assessed in adults by self-report questionnaires. Here, we use a short computer-based task in which participants experience several scripted scenes that offer opportunity to display inhibited behavior, to investigate how closely task behavior correlates with BI assessed via questionnaire, and if so, whether this correlation can be modified depending on whether participants are, or are not, specifically instructed to respond in a way that simulates how they normally behave.

1.1. The personality trait of behavioral inhibition (BI) and risk for psychiatric disorders

Several personality traits and behavioral patterns have been associated with risk for psychiatric disorders. For example, the trait of behavioral inhibition (BI) is defined as a tendency to withdraw from or avoid from novel social and non-social stimuli (Kagan, Reznick, & Snidman, 1987; Morgan, 2006). BI is believed to be one of the most stable temperamental characteristics, although not all children with high BI develop into high-BI adolescents and adults (Degnan & Fox, 2007). BI can be identified in childhood based on structured interview and/or observation of behavior of the child when confronted with unfamiliar people and objects (Kagan, Reznick, Clarke, Snidman, & Garcia-Coll, 1984). Inhibited temperament in childhood is a risk factor for future development of anxiety disorders (Biederman et al., 1993; Hirshfeld et al., 1992; Pérez-Edgar et al., 2010; Svihra & Katzman, 2004) and for post-traumatic stress disorder (PTSD; Fincham, Smit, Carey, Stein, & Seedat, 2008; Kashdan, Morina, & Priebe, 2009).

1.2. Limitations of questionnaire tools for assessing BI

In adults, BI is most often assessed through self-report questionnaires, which ask respondents to rate themselves on perceived levels of inhibition relative to implicit social norms. Tools specifically designed to assess BI and avoidance behavior include the Retrospective and Concurrent Self-Report of Inhibition (Reznick, Hegeman, Kaufman, Woods, & Jacobs, 1992), the BIS/BAS Scale (Carver & White, 1994), and the Adult and Retrospective Measures of Behavioural Inhibition (AMBI/RMBI; Gladstone & Parker, 2005; Gladstone, Parker, Mitchell, Wilhelm, & Malhi, 2005); other widely-used and well-validated questionnaire tools exist to assess broader concepts of state and trait anxiety, such as the Spielberger State-Trait Anxiety Inventory (STAI; Spielberger, 1983), the Generalized Anxiety Disorder Severity Scale (GADSS; Shear, Belnap, Mazumdar, Houck, & Rollman, 2006), and the Beck Anxiety Inventory (BAI; Beck, Epstein, Brown, & Steer, 1988). Many of these tools ask the respondent to rate his/her personality and feelings with respect to implicit social norms (e.g. “I am shy,” “I am nervous”), which are open to individual interpretation, particularly in participants who may not have accurate understanding of social norms.

The AMBI is one questionnaire tool that attempts to remediate this issue in part by asking the respondent to report on frequency of specific behaviors (e.g., “Do you tend to introduce yourself to new people?” “Do you prefer your own company over the company of others?”), rather than evaluating oneself to implicit social norms. AMBI scores have previously been shown to accurately predict anxiety vulnerability (Gladstone et al., 2005) and to correlate with PTSD symptoms (Myers, VanMeenen, McAuley, et al., 2012; Myers, VanMeenen, & Servatius, 2012).

AMBI and the other abovementioned questionnaires have proven useful in elucidating the construct of BI and its relation to risk of anxiety and PTSD. However, there are inherent limitations to the use of any self-report questionnaire. The most obvious limitation is the potential for response bias and demand characteristics (for a good recent review of these issues, see McCambridge, de Bruin, & Witton, 2012) . For example, some participants may (consciously or unconsciously) understate inhibited behavior in order to appear well-adjusted or conform to a positive view of self; others (particularly those with a diagnosed psychiatric disorder) may overstate inhibited behavior in order to conform to expected symptoms (e.g. conforming to an accepted sick role).

Beyond this, some populations (particularly those with psychiatric disorders) may not have particularly good insight into or recall of their own behavior. For example, individuals with major depressive disorder display mood-congruent memory effects, in that they are more likely to recall negatively- than positively-valenced information (Matt, Vázquez, &

Campbell, 1992). Individuals with PTSD may be similarly impaired at accurate self-report, due to a tendency to overgeneralize autobiographical memories (e.g., Moradi, Abdi, Fathi-Ashtiani, Dalgleish, & Jobson, 2012; Ono, Devilly, & Shum, 2015 May 11 [Epub ahead of print]); in fact, a recently-defined subtype of PTSD involves dissociative symptoms, including disruptions in memory, identity, and perceptions (Bennett, Modrowski, Kerig, & Chaplo, 2015 May 25 [Epub ahead of print]; Tsai, Armour, Southwick, & Pietrzak, 2015). Finally, the form factor of a paper-and-pencil questionnaire may not be ideal for use with populations that have limited attention span or poor vocabulary and/or literacy skills.

1.3. Use of interactive, computer-based tools to assess personality

A viable alternative to questionnaires may be interactive virtual environments, in which the user experiences simulated situations via an avatar, a graphical representation of the user whose behavior the user controls (Blascovich et al., 2002). Such environments are common in online computer gaming, and are increasingly used in social networking (Bente, Rüggenberg, Krämer, & Eschenburg, 2008) and educational platforms (e.g., Danforth, Procter, Heller, Chen, & Johnson, 2009; Foster, 2008). In many of these environments, avatars may be customizable in appearance, may be realistically animated, and may move through sophisticated simulated environments and interact with numerous other characters, some of whom are controlled by the computer and some controlled by other users playing the game.

A recent body of literature suggests that, although users tend to create avatars that look like a physically idealized self (e.g., Dunn & Guadagno, 2012), the differences are often surprisingly modest, with most avatars reflecting the user's real or perceived self (Bessière, Seay, & Kiesler, 2007; Ducheneaut, Wen, Yee, & Wadley, 2009; Vasalou & Joinson, 2009), although the similarity of avatar to perceived self may depend on the context in which that avatar is to be used (e.g., social networking vs. online gaming environments; Sung, Moon, Kang, & Lin, 2011).

To date, only a few intriguing studies have compared the personality of online avatars ("in-world" personality) to the real-life personality of the user ("out-world" personality), along the Big Five personality dimensions (Ducheneaut et al., 2009; John, Naumann, & Soto, 2008; McCreery, Krach, Schrader, & Boone, 2012). One study found that, while in-world personalities tend to be idealizations (e.g. slightly higher in Openness and Agreeableness, lower in Neuroticism, compared to out-world personalities), most in-world and out-world personalities are remarkably congruent (Ducheneaut et al., 2009). When participants are asked first to self-report their own personality across the Big Five dimensions and then to create an avatar and rate its personality, the ratings of actual self and avatar are significantly though imperfectly correlated (Sung et al., 2011).

These results suggest the possibility of leveraging virtual environments to probe a participant's personality traits; in effect, instead of using a questionnaire to ask, "How closely does this adjective describe you?" an avatar-based assessment can probe, "How would you normally act in a situation like this?"

1.4. Design of the current study

As a first step towards this goal, in the current study we created a simple computer-based task, in which the user deployed an on-screen character ("avatar") through several virtual scenarios representing novel social situations with defined choice points that might evoke inhibited behavior in participants prone to BI. Our primary outcome measure was how closely the participant's behavior at these choice points would correlate with the participant's level of BI, as self-assessed by a traditional paper-and-pencil questionnaire. For the latter, we used the AMBI, because the behavior-directed items of the AMBI thus lend themselves fairly naturally to translation for scripting in an online environment. In addition, prior studies with the AMBI suggest that it exhibits useful psychometric properties, including a broad (and approximately normal) distribution of scores obtained in samples of putatively healthy young adults (Caulfield, McAuley, & Servatius, 2013; Holloway, Trivedi, Myers, & Servatius, 2012; Sheynin et al., 2013).

In two experiments presented here, we collected data from fairly large samples of putatively healthy young adults (college students), to examine whether scores obtained from the computer-based task would correlate with BI scores obtained from the traditional paper-and-pencil questionnaire format. This leads to our central hypothesis:

> **H1.** Scores on the computer-based task correlate strongly with participants' self-assessed BI, indexed as AMBI scores.

1.4.1. Contrasting behavior when instructed to behave like "self" vs. like a hypothetical other—

While we anticipated that most participants would follow instructions in the game to act "as they would in real life," this is of course not the same thing as directly observing or measuring actual, real-world behavior. Within the scope of the current study, if participants indeed follow these instructions, resulting in a correlation between AMBI and inhibited behavior in the task, then we should be able to manipulate (weaken) this correlation by altering the instructions. To test this idea, in Experiment 2 we considered two groups of participants, one receiving the usual instructions to behave in a manner consistent with their actual self, and a second group receiving instructions to act like a hypothetical other ("a typical Rutgers student"), even if this is not how they themselves normally behave. This leads to our secondary hypothesis:

> **H2.** The correlation between task scores and AMBI scores should be weaker in participants instructed to behave like a hypothetical other ("a typical Rutgers student"), than in participants instructed to behave as they normally do in real life.

In addition, we considered the possibility that, when participants were asked to adopt the persona of a "typical" Rutgers student, they might behave differently depending on their own level of BI. For example, participants with high BI might consider the "typical" Rutgers student to be less inhibited than they themselves, or participants with low BI might consider their peers to be generally more inhibited than they themselves. Although this is a speculative idea, which falls short of an explicit hypothesis, it constitutes an ancillary research question which we investigated by secondary analysis:

RQ1. Do participants high or low in BI produce game behavior suggesting that they believe their peers are similar, or different, in BI from themselves?

1.4.2. Possible effects of gender and computer gaming experience—Although computer use – and computer gaming involving the use of avatars – is increasingly widespread, particularly among younger adults, it can by no means be assumed that all young adults are equally familiar with computers or computer games. It is possible that participants who are used to game play involving an on-screen avatar will more easily manifest realistic behavior on the current task, increasing the degree to which on-screen behaviors correlate with participant BI as assessed by AMBI; conversely, it is possible that participants who are used to deploying sophistic and customized avatars in highly complex and immersive RPGs might be paradoxically less likely to fully engage in the current task, due to the fairly simple animation and customization options, decreasing the degree to which their onscreen behaviors correlate with AMBI scores.

Additionally, although the number of female gamers is rising, computer game play has traditionally been dominated by males (Ogletree & Drake, 2007; Padilla-Walker, Nelson, Carroll, & Jensen, 2010; Winn & Heeter, 2009), and this is particularly true for RPGs which are often centered around violence or warfare. Thus, there may be sex differences in our sample relating to prior experience with computer avatars, and this could affect the degree to which males vs. females generate on-screen behavior consonant with their self-assessed BI. Additionally, some prior research suggests that females may reveal more information about themselves online than men do, while other research suggests that female players may be more likely than males to create idealized avatars or to engage in gender-bending in constructing avatars, particularly in environments perceived as male-centric (for review, see Dunn & Guadagno, 2012). These tendencies might lead female participants to produce behavior in the current task that conforms more, or less, closely to their real-world behavior, compared to male participants.

To better understand address these issues, we followed the computer task with a post-test questionnaire, to request details regarding participants' familiarity with computers, computer game play, "attachment" to avatar, and how closely participants chose responses in the game similar to how they would normally behave in such situations in the real world; we also recorded subject gender. We also asked participants simply to rate how closely their behavior in the game mimicked their normal behavior in the real world. This led to the following research questions:

RQ2. If there is a correlation between task and AMBI scores, is the strength of the correlation affected by whether participants regularly play computer-based games (particularly RPGs)?

RQ3. If there is a correlation between task and AMBI scores, is the strength of the correlation affected by participants' gender?

RQ4. If there is a correlation between task and AMBI scores, is the strength of the correlation affected how well participants "like" their chosen avatar?

RQ5. How closely do subjects report exhibiting avatar behavior similar to their own real-world behavior, within the task?

2. Experiment 1

2.1. Methods

2.1.1. Participants—114 Rutgers University-Newark undergraduate students (mean age 21.4 years, SD 5.0; 64.9% female) were recruited via a departmental subject pool, in which students sign up to participate in research studies in exchange for credit in a psychology class. Given the highly diverse student body at this campus, there was a corresponding diversity in participants' self-reported races and ethnicities, with 20 self-reporting race as Black, African, or African-American, 18 self-reporting as Asian, 28 self-reporting as White or Caucasian, and the remaining 48 self-reporting as Mixed-Race or other; 29 participants self-reported ethnicity as Hispanic. Because it seemed possible that completing a questionnaire to assess BI could affect subsequent behavior in the computer-based task, or vice versa, order of administration of questionnaire and task was counterbalanced across participants.

Written informed consent was obtained from all participants before initiation of any testing. Experimental procedures were approved by the Institutional Review Board at Rutgers University and conformed to guidelines established in the Declaration of Helsinki for protection of human subjects.

2.1.2. Questionnaire—All participants completed the paper-and-pencil AMBI questionnaire (Gladstone & Parker, 2005). The questionnaire contains 16 questions in which the participant is requested to report on current (adult) behavior when entering a new or unfamiliar social situation, or new and unfamiliar surroundings. Specific questions ask whether participants tend to become vigilant and wary, awkward and quiet; whether they typically approach and speak with people they don't know or withdraw and keep quiet; and whether they feel physically anxious in new situations. Other questions ask whether the participant generally prefers his/her own company to those of others, prefers large and lively social events or solitary leisure activities, and enjoys adventure holidays and activities with some element of risk. AMBI scores range from 0 to 32, with higher scores indicating higher levels of BI.

2.1.3. Computer-Based Task—The computer-based task was conducted on a Macintosh laptop, with software programmed in the SuperCard programming language (Solutions Etcetera, Pollock Pines, CA), and took about 10 minutes for participants to complete. Participants were tested individually, in a quiet room, with the participant seated at a comfortable viewing distance from the screen; the experimenter was present to read the instructions aloud and then withdrew but remained nearby.

At the start of the task, an instruction screen appeared: "Welcome to the game. First you will select a character to represent you as you play the game. Then, you will interact with other characters in the game. You'll be asked to make decisions and describe the way you feel. You should think of the character in the game as yourself and try to make the same decisions as you would in real life. Click the 'go on' button below to start the game." The participant then used the mouse to click on a button labeled "Go on" at the bottom of the screen. Another screen then appeared with eight possible characters ("avatars"), instructing the participant to click on one to select it, and then click "Go on" (Figure 1A). The characters were designed to be fairly neutral in appearance, clothing, and facial expression, but with some variability in hair and skin tone; four were male and four were female.

The task itself consisted of two scenarios, each with a script that included a number of choice points; at each choice point, a short text description was provided together with an image showing the character experiencing this event, and the participant was presented with three response options. The participant selected one by clicking on the box next to that option and then clicking the "Go on" button at the bottom of the screen. The script was designed so that the same sequence of events and response options appeared for all participants, regardless of their prior responses, although this lack of contingency should not have been obvious to the participant. The two scenarios (attending a party full of strangers, and helping out at a construction site) were designed to describe relatively unfamiliar situations that might be expected to evoke social anxiety or avoidance. Within each scenario, specific choice points were meant to parallel AMBI items, with a few additional choice points thrown in where responses were not scored but which served to keep the script moving forward while attempting to partially disguise the psychological constructs being probed.

The complete script, including possible responses, is provided in the Appendix. In brief, the first scenario described a party (14 choice points). At the first choice point, the avatar was pictured standing against a background of grass and sky and holding a cell phone (Figure 1B). The description read: "You're in the park when your cellphone rings. It's your cousin Frank, inviting you to a party. Frank's friends will be there, but you won't know anyone other than Frank. You…" Below this description appeared three options. One, "Say you'll attend – you like meeting new people," was intended as an uninhibited response; another, "Say no – you'd rather spend time alone doing something you enjoy," was intended as an inhibited response; the third, "Invite Frank over to watch a movie instead," was intended as intermediate between the two.

Regardless of the participant's response to the first question, the script continued with the avatar proceeding to the party, and experiencing a series of events that included arrival at the party earlier than Frank and having to decide whether to wait outside or go in alone (Figure 1C); meeting the host of the party; having to make casual conversation with a stranger; etc. In 12 of the 14 cases, the choice point was accompanied by three response options, intended to be uninhibited, inhibited, or intermediate; response options were presented in pseudorandom order on the screen. Two additional choice points were interposed that asked questions unrelated to behavioral inhibition (one asking about preferred party foods, and one asking about preferences in music); these responses were unscored.

The second scenario implemented a script in which the avatar was asked to help out at a construction site where a charity was helping to rebuild homes destroyed by a recent storm (similar to a Habitat for Humanity project). At each of 8 choice points, the participant was presented with three possible responses meant to probe for inhibited, uninhibited, or intermediate behavior (e.g. Figure 1D). A question at the start of the scenario explicitly asked whether this scenario was novel for the participant (i.e. whether the participant had participated in similar projects before), and an additional unscored question was interposed during the scenario asking about preferred lunch choices.

In total, then, there were 12 questions from the party scenario and 8 from the construction-site scenario meant to probe BI; for each, the participant's response was coded as 2 points for the inhibited option, 0 points for the uninhibited option, and 1 point for the intermediate option; thus, for each participant, a total score was generated that could range from 0 to 40 points; there were also subscores for the party (0–24 points) and construction (0–16 points) scenarios. The question regarding novelty of the construction-site scenario was coded separately to be used as a grouping variable in assessing responses related to this scenario depending on whether such a scenario was novel, partly-familiar or very familiar to the participant. The three questions related to food and music were not included in any scores.

Following completion of the computer-based task, participants were given a short post-test questionnaire which asked them (1) to rate their familiarity with computers on a four-point scale (1=NONE, 2=A LITTLE BIT, 3=MODERATE, 4=HIGH); (2) whether they ever played computer games (Yes/No) and if so, which games and how many hours in a typical week; (3) whether they had been able to find character which they liked to represent them (and if not, why not); (4) whether they cared about what happened to the character in the game (1=NOT AT ALL, 2=A LITTLE BIT, 3=SOMEWHAT, 4=A LOT); and (5) whether they made decisions in the game that were similar to the way they normally act in real life (1=NOT AT ALL, 2=A LITTLE BIT, 3=MOSTLY, 4=EXACTLY LIKE ME).

2.2. Results

2.2.1. Questionnaire results—Mean score on the AMBI was 14.5 (SD 5.7); on average, females scored slightly higher than males (Females: M=15.3, SD 6.1; Males: M=13.0, SD 4.5; $t(112)=2.05$, $p=0.042$). On the post-test questionnaire, all but 3 participants rated themselves as moderate or high in computer familiarity (questionnaire item #1, M=3.6, SD 0.6); 43 participants (37.7%) reported regularly playing computer games. Participants who reported playing computer games were then asked to specify game(s). Specific role-playing games (RPGs) mentioned by name included The Sims, League of Legends, and Minecraft; however, RPGs were also scored if the participant named or described any game involving player point-of-view and/or design and control of an on-screen character. Even given this fairly generous definition of RPGs, only 10 participants (8.8% of total sample) reported RPG play; the remaining gamers reported other types of games (e.g. card games, puzzles, word games, strategy games such as Candy Crush or Bejewelled) or failed to specify. There were no gender differences on computer familiarity, self-reported game-playing, or types of games played (all $p>0.100$).

2.2.2. Avatar selection and comparison to real life behaviors—At the start of the computer-based task, all participants selected same-sex avatars, although they had not been explicitly instructed to do so. In general, those participants reporting race as Black/African-

American or Mixed/Other, or reporting ethnicity as Hispanic, tended to choose darker-skinned avatars, while those self-reporting as White/Caucasian or Asian tended to choose fairer-skinned avatars. While 87% of participants reported on the post-test questionnaire that they had been able to find an avatar they liked, the remaining participants typically commented that they “could not find an avatar that looked like me” or that “avatars were too stereotypical.” While participants tended to report only modest levels of concern about what happened to the avatar (questionnaire item #4, M=2.7, SD 0.8), most participants reported high correspondence between their responses in the game and how they normally act in real life, with all participants selecting either “MOSTLY” or “EXACTLY” on questionnaire item #5 (M=3.4, SD 0.5). There were no gender differences on any of the avatar-related questionnaire items (all p>0.100).

2.2.3. Responses on the computer-based task—On the computer-based task, mean total score was 16.9 (SD 5.8), with subscores of 9.5 (SD 4.0) on the party-scenario questions and 7.4 (SD 2.3) on the construction-scenario questions. The total set of 20 scored questions showed good inter-item reliability, with Chronbach’s alpha of α=.778; considering the two scenarios separately, α=.718 for the 12 party-scenario questions but α=.477 for the 8 construction-scenario questions, although alpha is often deflated when a small number of items are considered.

Participants used the entire range of response options; thus, at least one participant endorsed each response option for each question. The grand mean of BI scores, averaged across participants and across questions was 0.84 (SD 0.75), which was not significantly different from the theoretical mean of 1.0 which would be expected if all participants had chosen the intermediate response to all questions (one-sample t-test, t(19)=1.43, p=.169).

2.2.4. Strong relation between BI as assessed by questionnaire and via task performance—Total scores on the computer-based task were strongly correlated with AMBI scores (Figure 2; r=.783, p<.001); this relationship remained unchanged after controlling for order of administration (task first or AMBI first, partial r=.783, p<.001). Both subscores were also significantly correlated with AMBI scores (Party scenario: r=.757, p<.001; Construction scenario: r=.634, p<.001).

Given that the construct of BI is associated with avoidance of or withdrawal from novelty, it was of particular interest to consider responses to the construction-scenario questions among those who did, or did not, report prior experience with such events. In fact, 49 participants endorsed the construction-scenario as “novel,” 47 as “partially-familiar,” and 6 as “very familiar;” there were no differences in distribution of these responses among males and females (χ^2=1.04, df=2, p=.594). Even after controlling for familiarity, construction-scenario subscores were still significantly correlated with AMBI scores (Figure 3A; partial r=.608, p<.001). However, as suggested by Figure 3A, both AMBI scores and task total scores were significantly lower in those who rated the construction scenario as “familiar” compared to the other groups (AMBI: F(2,113)=6.51, p=.002; task: F(2,113)=5.70, p=.004; in both cases, Tukey post-hoc tests showed that the “novel” group differed significantly from the “familiar” group, all p<.01, but that the “partially familiar” group did not differ from either of the other groups, both .04<p<.35).

Although total scores computed over all 20 questions on the computer-based task correlated strongly with AMBI scores, not all questions were equally predictive. Stepwise linear regression with the 20 scored task questions as predictors revealed that party question #1 was the single best predictor of AMBI scores (R=.579, R^2=.335, F(1,112)=56.58, p<.001), but this model could be significantly improved by considering additional questions. A model including 8 specific questions (party questions #1, 3, 4, 8, 14 and construction questions #2, 5, 7; see Appendix for question contents) provided the best prediction of AMBI scores (R=.820, R^2=.672, F(8,105)=26.89, p<.001), with the remaining 12 questions not contributing significantly to account for additional variance in AMBI scores (all p<.050). AMBI prediction based just on this subset of 8 questions was not significantly worse than AMBI predictions based on the entire set of 20 questions (R=.839, R^2=.704, F(19,94)=11.75, p<.001); R^2 of change=.032, p=.532).

As noted above, females scored higher on AMBI than males; on the computer-based task, the gender difference was again significant (total task score, Female: M=18.2, SD 6.1; Male: M=14.7, SD 4.4; t(112)=3.21, p=.002). This difference was primarily due to gender difference on the party-scenario questions (Female M=10.5, SD 4.1; Male M=7.7, SD 3.3, t(112)=3.69, p<.001) with no significant difference on the construction-scenario questions (Female M=7.7, SD 2.5; Male M=6.9, SD 2.0; t(112)=1.62, p=.109). Considering the distribution of inhibited, uninhibited, and intermediate responses to all 20 scored questions, only two questions met Bonferroni-corrected significance levels (α=.0025) for gender difference: these included a party-scenario question about initiating conversation with a stranger (Party #7), and a construction-scenario question about willingness to work in a dangerous physical location on the project (Construction #4); in both cases, females endorsed inhibited responses more often than males. However, gender differences on several additional questions reached uncorrected significance (.0025<p<.050), including two party-scenario questions about chatting to strangers and enjoying adventure holidays (Party #5 and #14), for which females endorsed the inhibited response more often than males, one construction-scenario question about preferring to work “in the center of the action” vs. staying out of the way (Volunteer #3), in which females endorsed an intermediate answer (“help out wherever help is needed most”) more often than males, and one construction-scenario question about feeling excited vs. reluctant about upcoming participation on the work event (Volunteer #2), for which females endorsed the uninhibited response more often than males.

Finally, we considered whether attachment to avatar and prior experience with computer games (particularly RPGs) might mediate the relationship between task scores and AMBI scores. In fact, the strong relationship between task and AMBI scores remained, after controlling for how strongly participants reported “liking” their avatar (Figure 3B; partial r=.782, p<.001) and for type of computer games played (Figure 3C; partial r=.781, p<.001). Although there was a limited range in how strongly participants endorsed having “acted like real life” in the computer-based task, there was no significant difference in task scores between those who endorsed MOSTLY vs. EXACTLY LIKE ME (t(112)=0.48, p=.628).

3. Experiment 2

Experiment 1 demonstrated that a brief computer task, in which participants guided an onscreen character ("avatar") through several scripted scenarios by selecting from a set of possible responses at each choice point, could predict the personality trait of BI, as assessed by self-report questionnaire, providing support for our primary hypothesis (H1). Experiment 2 was designed to replicate and extend these results, by assigning participants to one of two conditions. In the first (condition "U"), participants were asked to make the avatar behave "the way you normally do in real life;" in the second (condition "R"), participants were asked to make the avatar behave "like an average Rutgers student," even if this was not the way they themselves typically behave. Our secondary hypothesis (H2) was that results from condition U would replicate the strong correlation between task and AMBI scores observed in Experiment 1, but that results from condition R would show less (or even no) correlation between task scores and self-assessed BI.

3.1. Methods

212 participants were recruited from the Rutgers-University community and received class credits or $20/hour payment for participation in a single one-hour session. Two participants' data were subsequently lost due to computer failure, leaving 210 participants. The remaining sample (mean age 21.0 years, SD 5.2; 64.3% female) was again highly diverse, with 60 participants self-reporting as Black, African, or African-American, 39 as Asian, 43 as White or Caucasian, and 68 as Mixed-Race or other; 55 self-reported ethnicity as Hispanic.

Order of administration of AMBI and computer-based task was again counterbalanced across participants, as was assignment to experimental condition "U" (n=104) or "R" (n=106). Condition "U" was the same as Experiment 1, including instructions to "think of the character in the game as yourself and try to make the same decisions as you would in real life." Condition "R" was similar, except that the instructions asked the participant to "make the character behave the way that you think an average Rutgers student would behave – even if that's different from how you usually behave." Otherwise methods followed those of Experiment 1.

Data analysis was also similar to that from Experiment 1, except that experimental condition (U vs. R) was added as a factor where appropriate, and we also conducted a comparison of correlation (Fisher's r-to-z transform) and slope (determined from regression analysis) in the relationship between AMBI and task scores in the two conditions. In comparing questionnaire results between conditions, Mann-Whitney U-test rather than t-test was used where appropriate for analysis of ordinal data.

3.2. Results

In general, the results from Experiment 2, particularly condition U, replicate those observed in Experiment 1; full details are provided in Appendix 2. Here, we focus on the comparison between condition U and condition R.

Mean AMBI score was 14.9 (SD 5.0), similar to that observed in Experiment 1; again, females on average scored slightly higher than males, although the difference was not significant ($t(208)=1.31$, $p=.193$). On the computer-based task, total score averaged 17.6 (SD 5.9) in condition U and 16.1 (SD 5.1) in condition R, which was not significantly different (Welch's t-test used since Levene's test indicated significant deviation from assumption of normality; $t(202.01)=1.94$, $p=.053$). Subscores for the party and construction scenarios averaged 9.8 (SD 4.5) and 7.8 (SD 2.2) in condition U, and 8.8 (SD 3.8) and 7.3 (SD 2.2) in condition R; again, the differences between condition were not significant (all $p>.05$).

In condition U, total scores were strongly correlated with AMBI scores (Figure 4A; $r=.780$, $p<.001$); scores were also significantly correlated in condition R (Figure 4B; $r=.480$, $p<.001$), although the strength of the correlation was weaker in condition R than in condition U (Fisher's r-to-z transform, $z=3.73$, two-tailed $p<.001$). The slopes of the relationships were also different: for condition U, the slope (derived from linear regression) was 0.85 (y-intercept 4.7; SE of slope 0.068) while for condition R, the slope was 0.53 (y-intercept 8.4, SE slope 0.095). Student's t-test computed as the difference between the two slopes divided by the standard error of the difference between the slopes (square root of sum of squared error) revealed significantly steeper slope in condition U than condition R ($t=2.74$, $df=N-4=206$, $p=.007$).

As in Experiment 1, participants in group U reported high correspondence between their responses in the computer-based task and how they normally act in real life ($M=3.5$, SD 0.6); however, those in group R reported significantly lower correspondence ($M=2.9$, $SD=0.8$; $U=3207$, $p<.001$).

4. Discussion

In two studies, we examined the ability of a simple computer-based task to examine participants' decision to engage in virtual behaviors illustrating the personality trait of behavioral inhibition. Experiment 1 generally supported our central hypothesis (H1), in that task scores correlated strongly with BI as assessed by a validated self-report questionnaire; it provided no support for the ideas (RQ2-4) that the relationship was affected by potentially confounding variables such as participant gender, experience with computer games, and whether or not the participant "liked" his or her chosen avatar. Experiment 2 generally replicated these results and also provided support for our secondary hypothesis (H2) that the correlation between task performance and participant BI was significantly weaker, though not abolished, when participants were asked to perform like someone else (condition R) rather than to perform as they themselves would in real life (condition U). The fact that the correlation between task performance and BI was significantly weaker in condition R than condition U suggests that participants could modulate their performance, and that the instructions to perform "as you normally would" had an effect on behavior. Neither experiment provided support for the idea (RQ1) that participants either high or low in BI had a systematic bias to view the hypothetical "average Rutgers student" as more or less inhibited than themselves. However, all participants in Experiment 1, and those assigned to Condition U of Experiment 2, reported behaving "mostly" or "exactly" the way they normally act in real life (RQ5).

The possible implications of these findings, along with study limitations and future directions, are discussed further below.

4.1. Computer-based task to assess BI

To our knowledge, this study represents the first to examine the use of a computer-based task to assess the personality trait of BI, which has been linked to vulnerability to anxiety disorders (Hirshfeld et al., 1992; Svihra & Katzman, 2004) and PTSD (Fincham et al., 2008; Kashdan et al., 2009). However, some prior studies have used more sophisticated immersive environments to examine how online behavior relates to personality traits (Bessière et al., 2007; Fong & Mar, 2015; Wohn & Wash, 2013). For example, Wohn & Wash (2013) compared personality assessed via questionnaire with behavior in a virtual environment. They considered a game in which participants design a small town; results showed that external observers could make reasonably accurate judgments about participants' personality scores, based on snapshots of the completed town. One implication of this work is that, as in the current study, participant behavior can often bear a strong relationship to self-described personality. Further, in the Wohn & Wash study, observed personality (as inferred from game performance) was always closer to the participants' self-described "real" personality than their self-described "ideal" personality. However, as in the current study, the relationship between personality and performance was strong but not perfect (as would be indicated by Pearson's r=1). Wohn & Wash suggest several interpretations for this imperfect correlation, which also apply to the current study: specifically, it is possible that the observed behavior is a "true" representation of personality while questionnaire results are affected by demand characteristics such as social desirability bias; alternately, it is possible that the questionnaire results are highly accurate but that game behaviors reflect only some aspects of personality (or are imperfectly coded as personality metrics). Nevertheless, even though the correlations (both in Wohn & Wash, and also in the current study) are imperfect, they are highly significant, with large effect sizes, indicating that assessments of behavior in virtual environments can provide valid ways to assess aspects of personality.

Given that BI is an aspect of personality which confers risk for anxiety disorders and PTSD, a computerized tool might facilitate identifying those high in BI, and who are therefore at risk for developing such disorders. Such a tool could also be delivered online, allowing broad screening. In addition to possible use as a screening tool, it is also possible that virtual environments might be developed as therapeutic tools for those with pathological inhibition or avoidance, as a component of exposure therapy to help participants explore how they might change their behavior within the context of a virtual environment. Such environments might be completely under experimental control, as in the current study, or might involve true virtual worlds in which the participant interacts with other (real) individuals online. At least some studies suggest that some individuals – particularly those high in Introversion and/or Shyness – may be more able to express their "true self" to online friends than in face-to-face encounters (Amichai-Hamburger, Wainapel, & Fox, 2002; Marriott & Buchanan, 2014). Given that BI, introversion and shyness are highly overlapping constructs, it seems likely that individuals with BI may similarly find it easier to express their "true self" online. To our knowledge, this possible therapeutic use of online environments in anxiety and in anxiety-vulnerable individuals has yet to receive systematic study, although Pan et al. (2012)

showed that, after playing an immersive virtual reality game where the male participant is approached by a virtual women who initiates conversation, socially anxious (but not confident) males showed decreased social anxiety assessed via questionnaires. In another study, prosocial behavior in undergraduates was affected by prior game play in which they had been assigned prosocial (heroic) or antisocial (villainous) avatars (Yoon & Vargas, 2014). Thus, a possibility for future work would be to encourage inhibited participants to play in a “less inhibited” fashion than they would normally behave, to see if participants’ BI scores change after this play, relative to pre-play baselines. These results suggest that avatar-based tools might have some therapeutic utility by allowing participants to role-play uninhibited behavior within the relatively safe confine of a game environment.

An interesting and unexpected observation in Experiment 1 (Figure 3A) is the relationship between AMBI scores and self-reported familiarity with the construction-site scenario. Specifically, individuals self-reporting high degree of familiarity with the scenario had lower AMBI than individuals self-reporting that the scenario was completely novel. This relationship approached but fell short of significance in Experiment 2 ($p=.070$). Correlation cannot be used to confer causality, so the data from Experiment 1 cannot suggest whether individuals with low AMBI are more likely to seek out such volunteer opportunities, or whether participation in such activities might actually create a change in behavioral inhibition. Either possibility is interesting, and the latter in particular might suggest that volunteer activities with a social component might be considered in designing therapy for individuals with extremely high inhibition or social anxiety. However, this idea is beyond the scope of the current study.

4.2. Limitations and future directions

The current study has some limitations. First in contrast to some earlier studies, many of which considered highly immersive and interactive platforms, including commercial online multi-player environments, the current work used very simple two-dimensional “scenes,” without animation, and with a limited number of response options at each choice point. This design maximized experimental control and ability to translate responses directly into a score assessing inhibited behavior, but of course immersion was low. Indeed, a few participants spontaneously reported that the possible responses did not always represent the actions they wished to select, although this dissatisfaction appeared to revolve mainly around the non-scored “distractor” questions, such as which of several foods or musical styles were preferred. As such, this may not represent a serious limitation of the task in evoking inhibited and uninhibited behavior.

Several participants also reported dissatisfaction with the selection of avatars, particularly the limited number of hairstyles. In fact, prior work has suggested that hair (style and color) is the single most important customizable physical attribute for avatars (Ducheneaut et al., 2009), perhaps because it is a highly visible and therefore identifying attribute within a virtual world. Forcing participants to choose from a small set of avatars may introduce a confound, given prior studies suggesting that participants can alter their behavior based on features of an assigned avatar (Yee & Bailenson, 2007; Yoon & Vargas, 2014). However, even in commercial online platforms that have invested considerable resources into allowing users to customize hundreds of aspects of their avatar's appearance, from eyelash length to thigh circumference, users still often report dissatisfaction with the available options (Ducheneaut et al., 2009), and so this may not be a limitation that is easily remediated. On the other hand, even though avatar selection was limited in the current study, participants were still able to choose for themselves which avatar they wished to use in the task, and this may mitigate against some of the confound observed by Yee et al. when avatars were assigned.

In addition, the current study focused on a sample drawn from a population of young, high functioning, putatively healthy individuals, who would be expected to have high familiarity with computers. Although a relatively low percentage self-reported regular gaming, and even fewer specifically reported engaging in RPGs, it seems likely that the vast majority of participants were at least familiar with the basic concepts of online gaming (even if this was not a preferred recreational activity). Thus, an important extension to the current work will be to see if the same strong association between BI and task performance holds in an older sample, which may be somewhat less computer literate and have had somewhat less lifetime exposure to online games.

Although all participants in Experiment 1 and in Condition U of Experiment 2 reported behaving in the task as they would in real life (RQ5), with task behavior correlating strongly with participant BI, the task is not completely free of the limitations of self-report, since it requires participants to introspect regarding their own behavior and to accurately simulate this through their avatar. It is still possible that subjects (consciously or unconsciously) failed to accurately characterize their own behavior. Mitigating against this concern is the fact that participants explicitly instructed to behave like a hypothetical other (Condition R of Experiment 2) showed much weaker correlations between task behavior and their own AMBI, and also were less likely to report on the post-test questionnaire that the avatar's behavior matched their own. However, the concern remains; future research might begin to address this issue by comparing task performance against personality assessment or behavior report provided by a family member or close acquaintance.

An even stronger approach might be to compare virtual behavior, elicited in the computer-based task, against traditional data capture approaches that could reveal real-world behavior, or even more sophisticated virtual environments; for example, whereas the "avatar" in the current computer-based task is a simple cartoon figure that is placed into pictures representing scenarios, more sophisticated virtual environments allow the user to control a true avatar via his/her own body movements in three-dimensional space, often including haptic and other sensory feedback. In fact, many researchers believe that cognition fundamentally emerges as an interaction of the organism with the environment (Barsalou, Simmons, Barbey, & Wilson, 2003; Smith, 2005) . This aspect of cognition is not addressed by the current computer-based task, nor indeed by paper-and-pencil questionnaires such as the AMBI. Thus, an interesting avenue for future work could compare embodied cognition, including decisions whether to engage in or withdraw from novel social and non-social situations, against results obtained from less immersive approaches. Nevertheless, the current studies as well as existing literature suggest that even simple computer-based tasks (and indeed self-report questionnaires) can capture meaningful information, such as vulnerability to clinical disorders.

As such, the most important future extension of the current work will be to examine whether performance on this computer task correlates not only with the personality trait of BI, but also with degree of symptomatology in individuals with anxiety disorders or PTSD. If so, the task may be a useful screening tool as well as a marker of symptom severity. This could potentially be important in patient populations where demand characteristics are high and may compromise self-report questionnaire or even face-to-face interview; in addition, psychiatric populations often suffer from limited awareness of deficit, which can hinder the accuracy of questionnaires asking individuals to describe their own behavior. Future work is planned to assess this task in such populations.

5. Conclusions

The current work presents a short, computer-based task in which participants guide an “avatar” through a set of highly-controlled, onscreen scenarios designed to evoke behaviorally-inhibited or uninhibited behaviors. To our knowledge, this is the first investigation to examine how onscreen behaviors relate to BI, which is a personality trait specifically associated with vulnerability to anxiety and PTSD. Participants’ behavior on the task correlates strongly with behavioral inhibition, assessed by a validated self-report questionnaire (the AMBI). As such, this computer task may be a useful tool to screen for behavioral inhibition and anxiety risk, particularly in populations where self-report questionnaires may be problematic.

Acknowledgments

The authors wish to thank Nicole Anastasides, Nadine Gordon, and William R. Krusznis for helpful feedback on task design, and the Honors College at Rutgers University-Newark for providing testing space. This research was supported by funding from the Clinical Science Research and Development (CSR&D) Service of the VA Office of Research and Development (I01 CX000771), by the NSF/NIH Collaborative Research in Cognitive Neuroscience (CRCNS) program and NIAAA (R01 AA018737), and by the SMBI. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript; the contents do not necessarily reflect the views of the Department of Veterans Affairs or the U. S. Government.

Appendix 1

Script for the computer-based task. Accompanying illustration described in square brackets; points scored for each response shown in parentheses (2 points = inhibited response, 0 points = uninhibited response, 1 point = intermediate response, not scored = does not contribute to score).

Party 1 [avatar alone in park]: You’re in the park, when your cellphone rings. It’s your cousin Frank, inviting you to a party. Frank’s friends will be there, but you won’t know anyone other than Frank. You:

- Say you’ll attend – you like meeting new people. (0)
- Say no – you’d rather spend the time alone doing something you enjoy. (2)
- Invite Frank over to watch a movie instead. (1)

Party 2 [same scene]: Frank insists, and eventually you agree. How do you feel about going to the party?”

- Not thrilled… You'd rather stay home. (2)
- It might be fun. (1)
- You're excited and eager to go. (0)

Party 3 [avatar alone outside a house]: You arrive outside the house. From the street, you an hear the sounds of music and people laughing inside. Just then your phone rings. It's Frank, saying he'll be a little late. You realize you won't know anyone at the party until he gets there. You:

- Decide to wait on the sidewalk until Frank gets there. (2)
- Feel a little nervous, but decide to go in to the party alone. (1)
- Have no problem going in and introducing yourself. (0)

Party 4 [avatar and host in doorway]: As you hang up your phone, the door opens. It's the host of the party. Seeing you there, she welcomes you in. The first thing you feel is:

- Happy to have arrived at the party. (0)
- Not sure what to make of the situation yet. (1)
- Nervous – it's hard to walk into a room full of strangers. (2)

Party 5 [avatar alone at party]: The host chats politely to you for a few moments, then excuses herself to check on something in the kitchen. You look around and realize you don't know anyone else there. You:

- Find someone who looks friendly, walk over, and say hi. (0)
- Check out the food and drinks. (1)
- Call your cousin Frank back and ask when he's coming. (2)

Party 6 [avatar with Frank and Jaime at party]: Just then, Frank arrives at the party. He has brought his friend Jaime with him, and he introduces you two to each other. Then he offers to go get everyone some snacks. Do you want:

- Nachos. (not scored)
- Vegetable sticks and dip. (not scored)
- Nothing. You ate before you came. (not scored)

Party 7 [avatar with Jaime at party]: Frank goes off to get the food, leaving you and Jaime alone to chat. What do you do?

- You start telling Jaime all about yourself. (0)
- You are a bit shy and let Jaime lead the conversation. (1)
- Check your phone for messages. (2)

Party 8 [same scene]: Jaime mentions liking adventure vacations with an element of risk, like mountain climbing and scuba diving. You say:

- Me too! (0)
- A little excitement is good, but sometimes I just want to relax. (1)
- That sounds too risky for me! (2)

Party 9 [same scene]: Jaime asks you, "If you could do anything for your next vacation, what would it be?" You say:

- Stay at home and relax. (2)
- Go somewhere new but relaxing. (1)
- Try somewhere new and exciting. (0)

Party 10 [same scene]: Jaime comments that the music at the party is pretty loud. You say:

- Honestly, I would have preferred a smaller, quieter party. (2)
- It's fine. (1)

- I like the noise and the lively activity. (0)
Party 11 [same scene]: The conversation turns to music. Jaime says, “I always wanted to play in a band, but I just don’t have the talent.” You say: - You really enjoy making music as well as listening to it. (not scored) - You enjoy listening to music, but you don’t play or sing well yourself. (not scored) - Honestly, you’re not really all that interested in music. (not scored)
Party 12 [with Frank and Jaime at party]: At last, your cousin Frank returns. He hasn’t brought the snacks, but he has a drink in his hand. You: - Politely excuse yourself, then go meet new people. (0) - Check your cellphone while Frank and Jaime talk. (2) - Just chat to Frank and Jaime for the rest of the evening. (1)
Party 13 [same scene]: After a couple of hours at the party, you notice that your cousin Frank has gotten pretty drunk and is talking about driving himself home. What do you do? - Gently suggest to Frank that he’s too drunk to drive. (1) - Take Frank’s keys away. (0) - Let Frank make his own decisions. (2)
Party 14 [with Frank and Jaime outside the house]: Jaime, who hasn’t been drinking, offers to drive Frank home. You say goodnight, and decide to go home yourself. The next time Frank invites you to a party, you will: - Say yes, and try to meet more people at the party. (0) - Say yes, and try to spend more time with your cousin. (1) - Say no – you’d rather stay home and watch movies. (2)
Volunteer 1 [avatar alone in park]: The next morning, your neighbor asks you to volunteer for a local charity that is building houses for people who lost their home in a recent storm. Have you ever done anything like this before? - Yes, you often volunteer to help on projects like this. (familiar) - Maybe once or twice. (partially-familiar) - No, never. (novel)
Volunteer 2 [same scene]: Your neighbor convinces you that it’s a really worthwhile cause, and you agree to help. He tells you where the worksite is, and reminds you to wear casual clothes. Thinking about it, you feel: - Excited – you like trying new things and meeting new people. (0) - Unsure – you could go either way. (1) - Reluctant – you’d rather donate money instead. (2)
Volunteer 3 [avatar with foreman at construction site]: As you arrive at the work site, you see the skeleton of a partially built house. There are people working on the roof, some working inside, and others standing in front, talking. The foreman comes to greet you, and asks you what you would like to do. - Learn something and be in the center of the action. (0) - Help out wherever help is needed most. (1) - Prepare information packets and stay out of the way. (2)

Volunteer 4 [same scene]: The foreman checks his list again and says he really needs someone to hammer nails. You agree to take the job. He asks you where you would like to work: - Up on the roof. (0) - Safe on the ground floor. (2) - Doesn't matter. (1)
Volunteer 5 [same scene]: The foreman asks if you can work in the upstairs bedroom. You agree. He points out where you're supposed to go, and you find a spot to work. What do you do next? - Grab a hammer and start working independently. (2) - Ask the foreman for someone to help you get started. (1) - Introduce yourself to everyone before you begin working. (0)
Volunteer 6 [at construction site with co-worker]: You find a spot to work and start hammering in nails. After a while, a new person comes over and begins working right beside you. - You feel awkward and try to move away. (2) - You don't care – you just keep doing your job. (1) - You introduce yourself and offer to help. (0)
Volunteer 7 [same scene]: You work for a couple of hours. Then you notice that it's getting close to noon -- time for lunch. What do you decide to do? - Go get lunch by yourself at your favorite nearby spot. (2) - Follow the others and see what they're doing for lunch. (1) - Invite some workmates to get lunch as a group. (0)
Volunteer 8 [construction site picnic]: Just then, the foreman announces that lunch was donated for all the workers. You look at what's available. What do you pick? - Hamburger. (not scored) - Chicken salad sandwich. (not scored) - Green salad. (not scored)
Volunteer 9 [same scene]: During lunch, you chat a bit to the foreman. He brings up a rather controversial issue about the upcoming election, and his opinion is exactly the opposite of yours. - You let him keep talking – no point in fighting. (2) - You try to change the subject. (1) - You argue with him. (0)
Volunteer 10 [same scene]: After lunch, you think about what you'd like to do for the rest of the afternoon. - You're comfortable sticking with the same job you were given. (2) - You tell the foreman you want to try something new. (0) - You ask the foreman what he wants you to do next. (1)
Goodbye screen [at construction site with foreman]: At the end of the day, it's time for everyone to leave. The foreman says thank you – you've done a great job and the house is looking great!

Appendix 2

Additional results from Experiment 2.

Results reported in the main text for Experiment 2 focus on the comparison between condition U and condition R. Condition U, which replicated the conditions of Experiment 1, in general provided a replication of those results. Specific results are reported below.

A.2.1. Post-test questionnaire and avatar selection

On the post-test questionnaire, all but 8 participants in Experiment 2 rated themselves as moderate or high in computer familiarity (questionnaire item #1, M=3.5, SD 0.6), with no gender differences ($t(208)=1.56$, $p>.120$); 53 participants (25.2%) reported regularly playing computer games and only 14 (6.7%) reported playing any form of RPGs. Whereas 24 of 135 females reported playing computer games, 29 of 75 males reported playing computer games (Yates-corrected chi-square=10.07, $p=.002$); of those who reported playing games, 5 of 24 females and 9 of 29 males reported playing RPGs, which was not a significant difference (Yates-corrected chi-square=0.28, $p=.599$).

Information regarding avatar identity was lost for four participants due to experimenter error; since the remaining data were intact for these participants, they were not excluded from analysis of questionnaire data or task performance. Among those participants where avatar identity was available, the vast majority selected same-sex avatars, although three male participants and one female participant chose avatars of the opposite sex. In general, those participants reporting race as Black/African-American or Mixed/Other tended to choose darker-skinned avatars, while those self-reporting as White/Caucasian tended to choose fairer-skinned avatars, with those self-reporting as Asian or Hispanic tending to choose avatars with intermediate skin tones.

As in Experiment 1, most participants (86.2%) reported that they had been able to find an avatar they liked, with the remaining participants generally commenting that they could not find an avatar that looked sufficiently like themselves or with an appropriate hairstyle or dress code (e.g. man with a turban or woman with a headscarf). While participants tended to report only modest levels of concern about what happened to the avatar (questionnaire item #4, M=2.7, SD 0.9), this did not differ between participants assigned to conditions U and R (Mann-Whitney U test, U=5502, $p=.980$). There were no gender differences on any of the avatar-related questionnaire items (all $p>.100$).

A.2.2 Performance on task and relationship with BI as assessed by AMBI

On the computer-based task, mean total score across all participants was 16.8 (SD 5.6), with subscores of 9.3 (SD 4.2) on the party-scenario questions and 7.5 (SD 2.2) on the construction-scenario questions. The total set of 20 scored questions showed good inter-item reliability, with Cronbach's alpha of $\alpha=.748$; considering the two scenarios separately, $\alpha=.730$ for the 12 party-scenario questions but $\alpha=.407$ for the 8 construction-scenario questions, although alpha is often deflated when a small number of items are considered. Unlike the prior experiment, there was no difference in total scores or either scenario subscore in males vs. females, in either condition (all $p>.300$).

Relationship between AMBI scores and task total scores is reported in the main text. Party scenario subscores were significantly correlated with AMBI scores in both condition U (r=.

759, p<.001) and condition R (r=.413, p<.001); again, the correlation was significantly stronger in condition U than condition R (z=3.96, p<.001). Construction scenario subscores were also significantly correlated with AMBI in both conditions (condition U: r=.545, p<.001; condition R: r=.400, p<.001) but there was no difference between conditions (z=1.34, p=.180).

Similar to the prior experiment, 95 participants endorsed the construction scenario as “novel”, 66 as “partially-familiar,” and 49 as “very familiar. The distribution was similar among participants assigned to condition U and condition R (chi-square=4.25, p=.120), and between males and females (chi-square=0.20, p=.906). After controlling for familiarity, construction-scenario subscores were still significantly correlated with AMBI scores in both condition U (r=.512, p<.001) and condition R (r=.391, p<.001). As in the prior experiment, in condition U, construction-scenario subscores were significantly lower in those who rated the construction scenario as “familiar” compared to the other subgroups (F(2,101)=7.06, p=.001; Tukey’s HSD showed that scores in the “very familiar” group were significantly lower than the “novel” (p=.001) and “partially-familiar” groups (p=.035), which did not differ (p=.685). However, unlike the prior experiment, there was no difference in AMBI scores in condition U as a function of familiarity (F(2,101)=2.73, p=.070); in condition R, there were no significant differences in either AMBI or construction-scenario subscores as a function of familiarity with the scenario (all F<2.0, all p>.100).

The prior experiment suggested that a subset of the 20 task questions could predict AMBI scores about as well as the entire set of 20 scored questions. Applying the same subset to predict AMBI scores in condition U also produced good prediction (R=.731, R-squared=.534, F(8,95)=13.61, p<.001); however, significantly better prediction was obtained when the full set of questions was used (R-square of change=.155, p<.001). In condition R, the reduced subset of questions produced a weaker prediction (R=.481, R-square=.231, F(8,97)=3.65, p=.001) which was not significantly worse than AMBI predictions based on the entire set of 20 questions (R-square of change=.082, p=.608).

After controlling for computer experience, type of computer games played, how strongly participants “liked” their avatars, and how strongly participants endorsed having “acted like real life” in the game, correlations between AMBI and task total scores remained similar (Condition U: all .700<r<.800, all p<.001; Condition R: all .400<r<.500, all p<.001).

References Cited

Amichai-Hamburger Y, Wainapel G, Fox S. “On the Internet no one knows I’m an introvert”: Extroversion, neuroticism, and Internet interaction. CyberPsychology and Behavior. 2002; 5(2): 125–128. [PubMed: 12025878]

Barsalou LW, Simmons WK, Barbey AK, Wilson CD. Grounding conceptual knowledge in modality-specific systems. Trends in Cognitive Sciences. 2003; 7(2):84–91. [PubMed: 12584027]

Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: Psychometric properties. Journal of Consulting and Clinical Psychology. 1988; 56:893–897. [PubMed: 3204199]

Bennett DC, Modrowski CA, Kerig PK, Chaplo SD. Investigating the dissociative subtype of posttraumatic stress disorder in a sample of traumatized detained youth. Psychogical Trauma. May 25.2015 [Epub ahead of print]).

Bente G, Rüggenberg S, Krämer NC, Eschenburg F. Avatar-mediated networking: Increasing social presence and interpersonal trust in net-based collaborations. Human Communication Research. 2008; 34(2):287–318.

Bessière K, Seay AF, Kiesler S. The ideal elf: Identity exploration in World of Warcraft. CyberPsychology & Behavior. 2007; 10(4):530–535. [PubMed: 17711361]

Biederman J, Rosenbaum JF, Bolduc-Murphy EA, Faraone SV, Chaloff J, Hirshfeld DR, Kagan J. A 3-year follow-up of children with and without behavioral inhibition. Journal of the American Academy of Child and Adolescent Psychiatry. 1993; 32(4):814–821. [PubMed: 8340303]

Blascovich J, Loomis J, Beall A, Swinth K, Hoyt C, Bailenson JN. Immersive virtual environment technology as a methodological tool for social psychology. Psychological Inquiry. 2002; 13:103–124.

Carver CS, White TL. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. Journal of Personality and Social Psychology. 1994; 67:319–333.

Caulfield MD, McAuley JD, Servatius RJ. Facilitated acquisition of eyeblink conditioning in those vulnerable to anxiety disorders. Frontiers in Human Neuroscience. 2013; 7:348. [PubMed: 23847516]

Danforth D, Procter M, Heller R, Chen R, Johnson M. Development of virtual patient simulations for medical education. Journal of Virtual Worlds Research. 2009; 2(2)

Degnan KA, Fox NA. Behavioral inhibition and anxiety disorders: Multiple levels of a resilience process. Developmental Psychopathology. 2007; 19(3):729–746.

Ducheneaut, N.; Wen, M-H.; Yee, N.; Wadley, G. Body and mind: A study of avatar personalization in three virtual worlds. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computer Systems (CHI 2009); Boston, MA. 2009.

Dunn RA, Guadagno RE. My avatar and me -- Gender and personality predictors of avatar-self discrepancy. Computers in Human Behavior. 2012; 28:97–106.

Fincham D, Smit J, Carey P, Stein DJ, Seedat S. The relationship between behavioural inhibition, anxiety disorders, depression and CD4 counts in HIV-positive adults: A cross-sectional controlled study. AIDS Care. 2008; 20(10):1279–1283. [PubMed: 19012085]

Fong K, Mar RA. What does my avatar say about me? Inferring personality from avatars. Personality and Social Psychology Bulletin. 2015; 41(2):237–249. [PubMed: 25576173]

Foster A. Using Second Life as a platform for education: Professor Avatar. The Education Digest. 2008; 73:12–17.

Gladstone GL, Parker G. Measuring a behaviorally inhibited temperament style: Development and initial validation of new self-report measure. Psychiatry Research. 2005; 135:133–143. [PubMed: 15922458]

Gladstone GL, Parker GB, Mitchell PB, Wilhelm KA, Malhi GS. Relationship between self-reported childhood behavioral inhibition and lifetime anxiety disorders in a clinical sample. Depression and Anxiety. 2005; 22(3):103–113. [PubMed: 16149043]

Guitton MJ. Cross-modal compensation between name and visual aspect in socially active avatars. Computers in Human Behavior. 2010; 26(6)

Hirshfeld DR, Rosenbaum JF, Biederman J, Bolduc EA, Faraone SV, Snidman N, Kagan J. Stable behavioral inhibition and its association with anxiety disorder. Journal of the American Academy of Child and Adolescent Psychiatry. 1992; 31(1):103–111. [PubMed: 1537760]

Holloway JL, Trivedi P, Myers CE, Servatius RJ. Enhanced conditioned eyeblink response acquisition and proactive interference in anxiety vulnerable individuals. Frontiers in Behavioral Neuroscience. 2012; 6:76. [PubMed: 23162449]

John, OP.; Naumann, LP.; Soto, CJ. Paradigm shift to the integrative Big Five trait taxonomy: History measurement, and conceptual issues. In: John, OP.; Robins, RW.; Pervin, LA., editors. Handbook of personality: Theory and Research. New York: Guilford; 2008. p. 114-158.

Joinson AN, Dietz-Uhler B. Explanations for the perpetration of and reactions to deception in a virtual community. Social Science Computer Review. 2002; 20(3):275–289.

Kagan J, Reznick JS, Clarke C, Snidman N, Garcia-Coll C. Behavioral inhibition to the unfamiliar. Child Development. 1984; 55:2212–2225.

Kagan J, Reznick JS, Snidman N. The physiology and psychology of behavioral inhibition in children. Child Development. 1987; 58(6):1459–1473. [PubMed: 3691195]

Kashdan TB, Morina N, Priebe S. Post-traumatic stress disorder, social anxiety disorder, and depression in survivors of the Kosovo War: Experiential avoidance as a contributor to distress and quality of life. Journal of Anxiety Disorders. 2009; 23(2):185–196. [PubMed: 18676121]

Marriott TC, Buchanan T. The true self online: Personality correlates of preference for self-expression online, and observer ratings of personality online and offline. Computers in Human Behavior. 2014; 32:171–177.

Matt GE, Vázquez C, Campbell WK. Mood-congruent recall of affectively toned stimuli: A meta-analytic review. Clinical Psychology Review. 1992; 12(2):227–255.

McCambridge J, de Bruin M, Witton J. The effects of demand characteristics on research participant behaviours in non-laboratory settings: A systematic review. PLoS One. 2012; 7(6):e39116. [PubMed: 22723942]

McCreery MP, Krach SK, Schrader PG, Boone R. Defining the virtual self: Personality, behavior, and the psychology of embodiment. Computers in Human Behavior. 2012; 28:976–983.

Moradi AR, Abdi A, Fathi-Ashtiani A, Dalgleish T, Jobson L. Overgeneral autobiographical memory recollection in Iranian combat veterans with posttraumatic stress disorder. Behavioral Research and Therapy. 2012; 50(6):435–441.

Morgan BE. Behavioral inhibition: a neurobiological perspective. Current Psychiatry Reports. 2006; 8(4):270–278. [PubMed: 16879790]

Myers CE, VanMeenen K, McAuley JD, Beck KD, Pang KCH, Servatius RJ. Facilitated acquisition of eyeblink conditioning in veterans with high behavioral inhibition, a risk factor for post-traumatic stress disorder (PTSD). Stress. 2012; 15(1):31–44. [PubMed: 21790343]

Myers CE, VanMeenen KM, Servatius RJ. Behavioral inhibition and PTSD symptoms in veterans. Psychiatry Research. 2012; 196:271–276. [PubMed: 22397911]

Ogletree S, Drake R. College students' video game participation and perceptions: Gender differences and implications. Sex Roles. 2007; 56(7/8):537–542.

Ono M, Devilly GJ, Shum DH. A meta-analytic review of overgeneral memory: The role of trauma history, mood, and the presence of posttraumatic stress disorder. Psychological Trauma. May 11.2015 [Epub ahead of print]).

Padilla-Walker L, Nelson L, Carroll J, Jensen A. More than just a game: Video game and Internet use during emerging adulthood. Journal of Youth and Adolescence. 2010; 39(2):103–113. [PubMed: 20084557]

Pan X, Gillies M, Barker C, Clark DM, Slater M. Socially anxious and confident men interact with a forward virtual woman: An experimental study. PLoS One. 2012; 7(4):e32931. [PubMed: 22509251]

Pérez-Edgar K, Bar-Haim Y, McDermott JM, Chronis-Tuscano A, Pine DS, Fox NA. Attention biases to threat and behavioral inhibition in early childhood shape adolescent social withdrawal. Emotion. 2010; 10(3):349–357. [PubMed: 20515224]

Reznick JS, Hegeman IN, Kaufman ER, Woods SW, Jacobs M. Retrospective and concurrent self-report of behavioural inhibition, and their relation to adult mental health. Development and Psychopathology. 1992; 4:301–321.

Shear MK, Belnap BH, Mazumdar S, Houck P, Rollman B. Generalized Anxiety Disorder Severity Scale (GADSS): A preliminary validation study. Depresssion and Anxiety. 2006; 23:77–82.

Sheynin J, Shikari S, Gluck MA, Moustafa AA, Servatius RJ, Myers CE. Enhanced avoidance learning in behaviorally-inhibited young men and women. Stress. 2013; 16(3):289–299. [PubMed: 23101990]

Smith LB. Cognition as a dynamic system: Principles from embodiment. Developmental Review. 2005; 25(3–4):278–298.

Spielberger, CD. Manual for the State-Trait Anxiety Inventory, STAI (Form Y). Consulting Psychologists Press, Inc; 1983.

Sung Y, Moon JH, Kang M, Lin J-S. Actual self vs. avatar self: The effect of online social situation on self-expression. Journal of Virtual Worlds Research. 2011; 4(1)

Svihra M, Katzman MA. Behavioural inhibition: A predictor of anxiety. Paediatrics and Child Health. 2004; 9(8):547–550. [PubMed: 19680482]

Tsai J, Armour C, Southwick SM, Pietrzak RH. Dissociative subtype of DSM-5 posttraumatic stress disorder in U.S. veterans. Journal of Psychiatric Research. 2015; 66–67:67–74.

Vasalou A, Joinson AN. Me, myself and I: The role of interactional context on self-preservation through avatars. Computers in Human Behavior. 2009; 25:510–520.

Winn J, Heeter C. Gaming, gender, and time: Who makes time to play? Sex Roles. 2009; 61(1/2):1–13.

Wohn DY, Wash R. A virtual "room" with a cue: Detecting personality through spatial customization in a city simulation game. Computers in Human Behavior. 2013; 29(1):155–159.

Yee N, Bailenson J. The Proteus Effect: The effect of transformed self-representation on behavior. Human Communication Research. 2007; 33:271–290.

Yoon G, Vargas PT. Know thy avatar: The unintended effect of virtual-self representation on behavior. Psychological Science. 2014; 25(4):1043–1045. [PubMed: 24501111]

Highlights

- The trait of behavioral inhibition (BI) can confer risk for anxiety disorders.
- Here, we developed a simple behavioral task to probe BI.
- Subjects guided an avatar through onscreen scenarios involving social interaction.
- Task performance correlated strongly with BI score assessed via questionnaire.
- Virtual environments may offer a useful format for personality assessment.

Figure 1.

Example screenshots from the computer-based task. (A) Participants are first asked to choose a character (“avatar”) to represent them in the game; here, the male figure at upper right is selected. (B) In the first question, the avatar is invited to a party where s/he will not know the people present; the participant chooses from three options including an uninhibited response (here, top option), an inhibited response (here, middle option), and an intermediate response (here, bottom option). (C) The party scenario proceeds regardless of the specific response to any question; here, the avatar arrives outside the party and chooses an inhibited (here, top), intermediate (middle), or uninhibited (bottom) action regarding entering an unfamiliar social environment. (D) Following the party-scenario questions, a second series of questions concern participation in a construction-site activity; again response options describe behaviors that are inhibited, uninhibited, or intermediate between the two extremes.

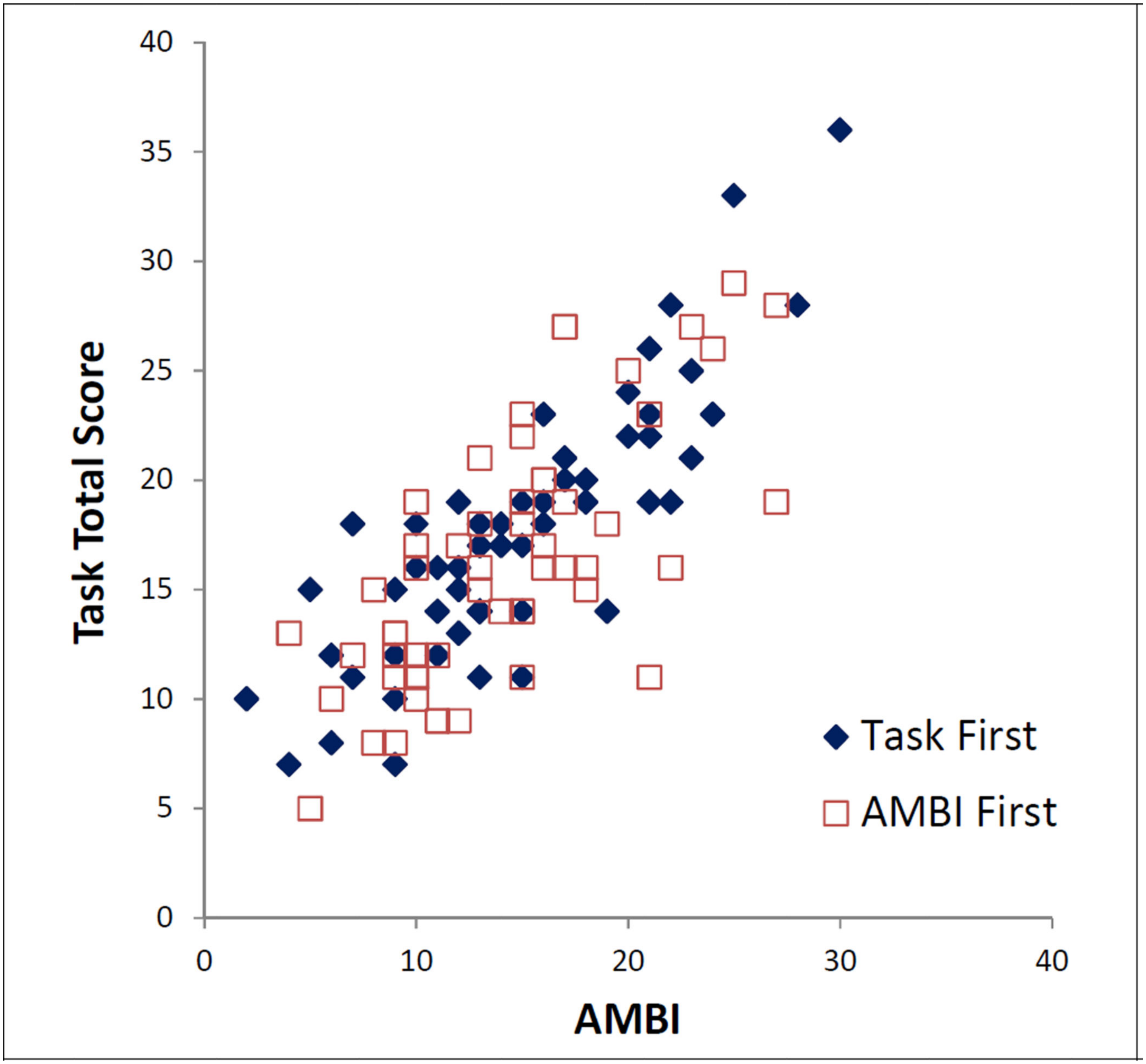


Figure 2.
Total scores on the computer-based task were strongly correlated with AMBI scores (r=.783, p<.001). This correlation remained unchanged after controlling for order of task administration (open s quares=computer-based task first, solid dots=AMBI questionnaire first; partial r=.783, p<.001).

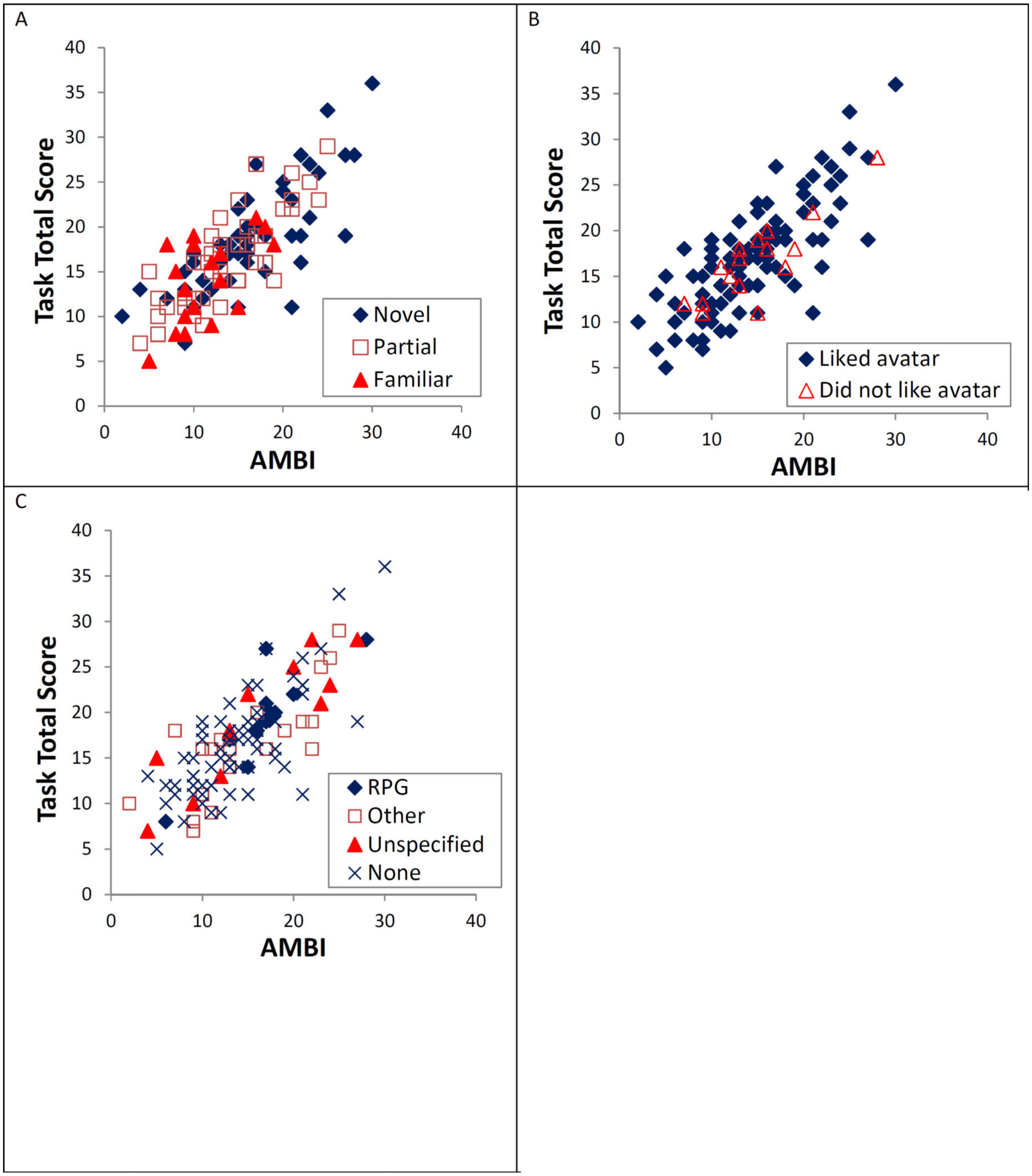


Figure 3.

Relationship between AMBI scores and task scores, controlling for other potentially confounding variables. (A) After controlling for familiarity with similar events, task scores were still significantly correlated with AMBI scores (partial r=.608, p<.001). Data points are coded according to whether participant reported that the construction scenario was novel (never participated in similar activities before), partially familiar (once or twice), or very familiar (participated many times in similar activities). Participants who reported the scenario as “familiar” had significantly lower AMBI and task scores than those who rated it as “novel.” (B) Correlation between task and AMBI scores remained after controlling for whether or not the participant reported finding an avatar s/he liked (partial r=.782, p<.001). (C) Correlation also remained (partial r=.781, p<.001) after controlling for whether participants reported playing videogames, and if so, whether these were primarily role-playing games (RPGs), non-role-playing games such as strategy or card games (Other), or did not specify type of games (Unspecified); “None”=participant reported not playing computer games.

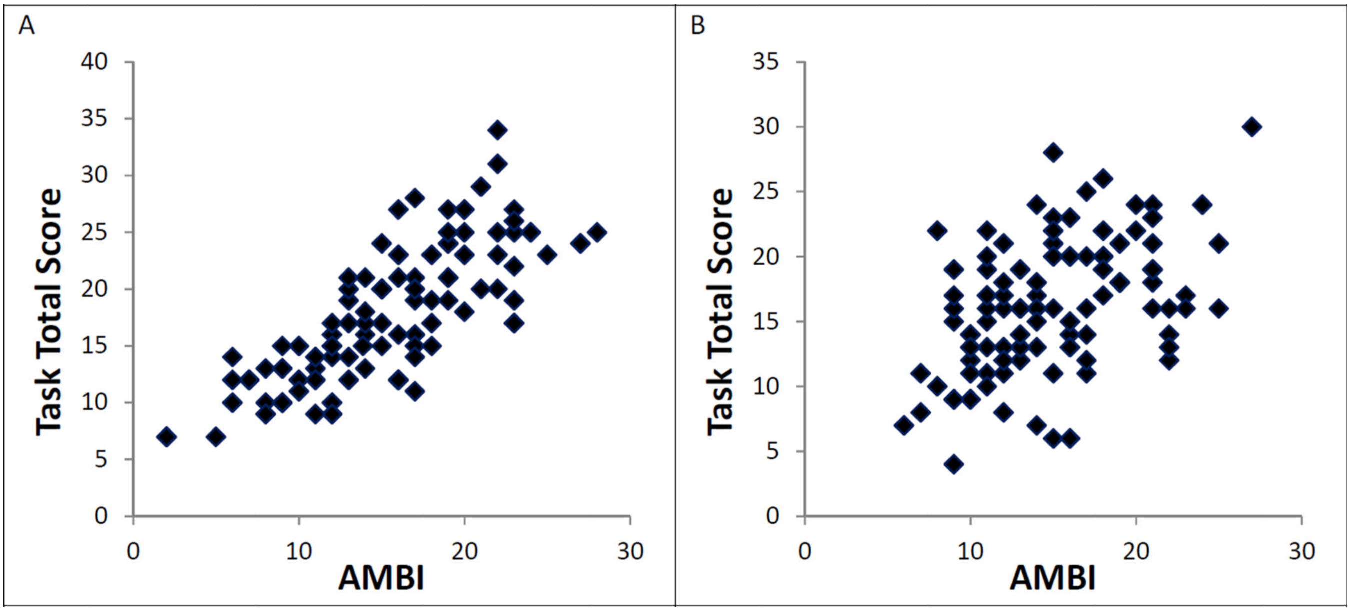


Figure 4.

Results from Experiment 2. (A) Condition U, which replicated Experiment 1, again showed a strong positive correlation between AMBI score and total score on the computer-based task (r=.780, p<.001). (B) Condition R, in which participants were asked to respond the way they thought an "average" Rutgers student would behave, even if it was different from how they themselves normally behave, showed a weaker but still significant correlation between task performance and BI assessed via AMBI (r=.480, p<.001).