Supplemental Materials

Materials and Methods

Photographs of targets. The lead author photographed each of the four confederates (i.e., *targets*) in front of a neutral background in the same lighting conditions. Two color photographs were taken of each target: one in which the target was instructed to smile and one in which she was instructed to display a neutral expression (see Fig. S1 for the photographs). All targets were undergraduates working as research assistants in a university in the Northeast of the U.S. and were Caucasian and female.

Time-1. The participant pool consisted of individuals (ages 18-30 years old) affiliated with a university in the Northeast of the U.S., who received an email invitation to participate in an online recruitment survey to be entered into a raffle for a $25 Amazon gift card. Eight hundred fifty-five individuals agreed to participate. Of these individuals, 636 (417 females, 217 males, 2 unspecified) completed the measures of interest for the current study and agreed to receive invitations for future studies (Instructions and items for the measures of interest can be found in the Materials posted in the Open Science Framework database).

As indicated in the main text, perceivers viewed neutral and smiling photographs of four targets. After viewing each photograph, perceivers were asked to evaluate the person in the photograph on four traits (*trustworthy*, *likeable*, *competent*, *aggressive*) and four items assessing willingness to engage in future interactions with the person (e.g., "*This seems like the kind of person whom I would like to get to know*."). After reverse scoring aggressiveness, these eight items were averaged to index liking of the target. Across the four targets, for the neutral photographs, Cronbach's α ranged from .86 to .88, *M*s ranged from 4.19 to 4.62, and *SD*'s ranged from .87 to .92 and for the smiling photographs, *α*s ranged from .86 to .89, *M*s ranged

from 4.62 to 5.09, and *SD*s ranged from .81 to .90. Perceivers also rated the person in each photograph on attractiveness (1 = *Not at all* to 7 = *Extremely*). Finally, they indicated their impressions of each target's personality on items measuring five major personality dimensions—extraversion, agreeableness, neuroticism, conscientiousness, and openness to experience.

Time-2.

Selection of Perceivers for the Lab Session. Out of the 636 perceivers who provided photograph-based judgments, we invited 348 who satisfied predetermined criteria to the lab session. The criteria were established to help minimize *perceiver effects* (e.g., the tendency for perceivers to either judge all persons favorably or unfavorably) and *target effects* (i.e., perceivers evaluating some targets more favorably than others) (Kenny & Albright, 1987), and were as follows: 1) for at least one of the targets, perceiver's impressions (mean of liking and attractiveness items) based both on the smiling and neutral photographs fell in the upper or lower quartile (i.e., a perceiver was *not* selected if his/her impressions based on one emotional display fell in the upper (lower) quartile but his/her impressions did not fall into the upper (lower) quartile based on the other emotional display.), 2) for at least one of the remaining three targets, perceiver's impressions did not fall in the upper (or lower) quartile, and 3) of the perceivers who were invited to interact with one of the four targets, roughly half expressed positive photograph-based evaluations, and the other half expressed negative photograph-based evaluations. The first criterion identified perceivers who either had favorable or unfavorable photograph-based impressions of the target regardless of the target's emotional display (smiling or neutral). Making sure that perceivers' judgments were similar across smiling and neutral facial expressions helped us obtain a more context-independent evaluation of the target (Todorov & Porter, 2014). The second criterion identified only those perceivers who discriminated among different targets in

their evaluations, which would serve to minimize perceiver effects. The third criterion ensured that each target interacted with roughly equal number of perceivers who held favorable vs. unfavorable photograph-based impressions of her, which would help minimize target effects. Finally, to help minimize simple context effects (e.g., in-person contact enhancing liking for all perceiver-target pairs by increasing familiarity) (Reis, Maniaci, Caprariello, Eastwick, & Finkel, 2011) we used the rank ordering within each target as our estimate of consistency.

Of the 348 perceivers invited to the lab session for monetary compensation ($20), 56 accepted the invitation. Perceivers who attended the lab session (vs. not) did not significantly differ in their gender composition (percentage male=%32 vs. %38, respectively, $\chi^2(2)=.92$, $p=.63$). Participants who attended ($M=19.91$) were younger than those who did not attend ($M=20.92$, $t(346)=2.67$, $p=.008$). Critically, participants who attended (vs. not) did not differ in photograph-based liking of the targets (all ps<.56). Nor did they differ in photograph-based ratings of targets on extraversion (all ps<.12), agreeableness (all ps<.13), neuroticism (all ps<.14), and openness (all ps<.56). Based on her photograph, one of the targets (Confederate 3) was rated as less conscientious by perceivers who attended ($M=4.56$) than those who did not ($M=4.87$, $t(341)=2.18$, $p=.03$). However, participants who attended (vs. not) did not differ in photograph-based ratings of the other targets on conscientiousness (all ps<.37). These results reveal that the two samples were largely similar in key variables of interest.

The lab session was held at least one month (range=35-212 days) after perceivers made the Time-1 photograph-based judgments. (The time between rating the photographs and completing the lab session did not moderate the photograph-live association, all ts<1). One perceiver was excluded from the final analyses because she expressed after the lab session that she already knew the target. For the remaining 55 perceivers (37 female, mean age=19.93 years,

SD=1.94 years), we obtained measures of liking, attractiveness, and personality based on photographs, by averaging across ratings of neutral and smiling photographs of each target (liking: *r*=.90, extraversion: *r*=.39, agreeableness: *r*=.27, neuroticism: *r*=.36, conscientiousness: *r*=.22, openness: *r*=.41).

Live Interaction Protocol. Our selection criteria resulted in 29 perceivers holding favorable photograph-based impressions of their interaction partner (*M*=5.60, *SD*=.24) and 26 perceivers holding unfavorable photograph-based impressions (*M*=3.50, *SD*=.44, *t*(53)=22.14, *p*<.001; see Table S2 for means and standard deviations for each target). Prior to the lab session, the experimenter checked the target's physical appearance to make sure that her makeup and hair were similar to those in her photograph. Targets were instructed to respond naturally to the perceiver during the interaction. Targets knew that the research focused on impressions based on photographs and those following live interactions, but were not told that we aimed to look at the association between the two. But, critically, targets were unaware that perceivers had been selected for the study based on their initial impressions of them and thus also were unaware whether perceivers held favorable or unfavorable impressions of them. To minimize repetitiveness of live interactions for targets and to minimize potential carryover effects from previous sessions, the study coordinator was asked to schedule the sessions such that each target completed only one session on a given day. This was possible for all but 4 of the 55 sessions due to scheduling constraints.

The first interaction was a 10-minute "trivia game" used in past work (Letzring, Wells, & Funder, 2006). The trivia game consisted of twenty questions of low, moderate, and high difficulty taken from a database of questions normed for difficulty (Nelson & Narens, 1980).[1] To minimize the repetitiveness of the trivia game we used 15 different sets of questions. The

study coordinator kept track of which sets targets had worked on in previous sessions and made sure that they did not complete the same trivia questions twice.[2] The perceiver and the target were given the following instructions: *"In this part of the experiment you two will be playing a trivia game together. One of you should read aloud the first question and then you should discuss what the answer might be. Even if you don't know the answer, take a few minutes to brainstorm about the question and what the answer might be. When you agree on an answer, write it down on this answer sheet. When you are done, go on to the next question. You can skip a question and later come back to it if you cannot come up with an answer after brainstorming. You will have 10 minutes to work on the questions. Please continue discussing the questions until your time is up."* At the end of the interaction, the perceiver and the target were given the following instructions: "*Now, we ask you to fill out a questionnaire about your experience. You will complete this questionnaire in separate rooms to maintain your privacy. Please be assured that all your responses will be confidential*." Then, the target left the room and perceivers and targets provided impressions of each other based on the interaction using the same liking (α=.89), attractiveness, and personality measures as those administered at Time-1. In addition, perceivers rated their own personality based on how they are feeling about themselves at the moment. Additionally, participants completed three items assessing their interaction partner's enjoyment of the interaction ("*How much do you think your partner enjoyed the interaction?*", "*How engaged was your partner during the interaction?*", "*How much during the interaction did your partner elaborate on his/her thoughts about the topic of conversation?*", 1 = *Not at all* to 7 = *Very much*). These three items were averaged to create an index of the perceivers' evaluation of the target's enjoyment of the interaction (α = .74). Finally, they completed a single item

assessing their own enjoyment of the interaction ("*How much did YOU enjoy the interaction?*", 1 = *Not at all* to 7 = *Very much*).

Then, the perceiver-target pair completed another 10-minute interaction used in past work (Letzring et al., 2006) in which they were given the following instructions "*Your task for the next 10 minutes is to get to know each other as well as possible and to learn as much as you can about what type of person each of you is*" ("getting-to-know interaction"). No other instructions were provided. At the end of this interaction, the perceiver and the target were given the same instructions as those given following the trivia game (described above). Then, the target again left the room and perceivers provided impressions of the target on the same measures (α= .89 for liking and .77 for perceived enjoyment of the target) and rated their own personality.

Following each interaction, the target also indicated liking (α=.91 for both interactions) and attractiveness of the perceiver, and rated the perceiver's and their own personality. They also completed the same three items as perceivers assessing their perception of the perceiver's enjoyment of the interaction (αs=.82 and .85 for the trivia and the getting-to-know interactions, respectively) and a single item assessing their own enjoyment. Means and standard deviations for all judgments across the two interactions are provided in Table S3. (All measures collected pertaining to interaction-based judgments are reported in the main text and/or supplemental materials. At the end of the study, perceivers also completed individual difference measures that are not relevant to the purposes of the present study and hence will not be discussed further.)

To ensure that perceivers were not aware that they had already provided photograph-based judgments of the target at Time-1, we used a funnel debriefing procedure. Specifically, we asked perceivers to indicate anything they found strange or unusual about the study, then to

guess the study hypotheses, and finally to report whether they were suspicious at any point in the study that their interaction partner might have been a confederate.

Additional Information about Coding Thin Slices of Behavior. Eight undergraduate research assistants (6 females) working in a university in the Northeast of the U.S. coded thin slices of the perceiver and the target's behavior obtained from the videos in terms of the person's *warmth ("How warm was this person being toward their interaction partner?", "How much do you think this person enjoyed the interaction?", "How engaged was this person during the interaction?"* , 1 = *Not at all* to 7 = *Very much).*[3] The coders viewed pre-specified segments of the videos using ELAN (Lausberg & Sloetjes, 2009), a free video annotation software which may be downloaded from the website of Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands (http://tla.mpi.nl/tools/tla-tools/elan). The coders rated all perceiver videos first, followed by all target videos. Before watching videos for the first time, the coders were given information about the nature of the interactions. Specifically, they were told that the first interaction was a 10-minute "trivia game" and the second interaction was a 10-minute "getting-to-know interaction." The coders also read the instructions delivered to the perceivers and the targets before each interaction (see Materials posted in the Open Science Framework database for the information provided to coders about the videos). Critically, the coders were blind to perceivers' photograph-based liking and the perceiver-target pairings. Agreement across coders was high (αs=.92 and .90 for the perceiver and the target videos, respectively). The behavioral warmth index obtained by averaging across warmth, enjoyment, and engagement ratings was highly reliable (αs=.97 and .97 for the perceiver and the target videos, respectively).

Data Analytic Strategy

Because perceivers' judgments were nested within targets, we used linear mixed models (LMMs) to account for interdependency among data points.

Estimating Statistical Power. Statistical power was estimated using the procedures outlined by Snijders and Bosker to estimate power in two-level mixed models using the software PINT v. 2.1 (Snijders & Bosker, 1993). Assuming residual perceiver-level variance of 50%, and random intercept variance of 10%, the standard error for the association between photograph- and interaction-based judgments was estimated as 0.09. Based on this estimate, at α=0.05, the power to detect a standardized association of 0.3 between photograph- and interaction-based judgments was 91%.

We also estimated the power to detect a difference of 0.3 in photograph-live association across the two interactions (trivia vs. getting-to-know). The test of whether the photograph-live association differs across interactions involves testing the difference between two dependent correlations (i.e., $r_{\text{photograph, trivia}} - r_{\text{photograph, getting-to-know}}$). Assuming $r_{\text{photograph, trivia}}$=.5, $r_{\text{photograph, getting-to-know}}$=.2, and $r_{\text{trivia, getting-to-know}}$=.6, the required sample size to detect the difference with 80% is 58 participants. The current sample size, N=55, was close to this estimate and provided 78% power.

Testing for Fixed vs. Random Intercepts in LMMs. For all dependent variables, we tested whether the intercept varied randomly across perceivers using a Log Likelihood test (Hayes, 2006). The intercept was allowed to vary randomly if the test result was statistically significant (i.e., χ^2>3.84). As a result, the photograph-live association in liking, openness, and conscientiousness were estimated using a fixed intercept and the photograph-live association in extraversion, neuroticism, and agreeableness were estimated using a random intercept. We also

tested whether the slopes representing the photograph-live association varied across perceivers using the same test. There was no evidence that the slopes representing the photograph-live association varied randomly (all χ^2s<2.42).

Comparing Photograph-Live Associations. To compare the size of the photograph-live associations across type of judgment (liking vs. personality), we constructed 95% confidence intervals around the difference of the two critical associations (e.g., photograph-live association in liking minus photograph-live association in personality) using the modified asymptotic method (Zou, 2007). This approach is preferred over the simple asymptotic approach to compare correlations based on simulation studies showing that it performs better even in small samples (e.g., *N*s around 50) both in terms of overall CI coverage and distribution of tail errors. We used a similar approach to test whether consistency in liking and personality judgments differed across interaction contexts (trivia vs. getting-to-know) except that this time one variable (photograph-based judgment) was common to both correlations. Given that both fixed-intercept and random-intercept models produced significant photograph-live associations in neuroticism and agreeableness, and that the methods for path analyses (see supplemental analyses below) and comparison of correlation coefficients are better developed for fixed-effects models, for simplicity, comparisons of the photograph-live associations across judgments (liking vs. personality) and interactions (trivia vs. getting-to-know) and the path analyses were performed by treating the target as a fixed factor.

Estimating Effect Sizes. In all LMMs, variance explained by a predictor was estimated by computing the proportional decrease in the residual variance of the model as a result of adding the predictor into the model.

Supplemental Analyses: Repeating the mediational models for behavioral confirmation using bootstrapping

Taylor, MacKinnon, and Tein (2008) recommended two approaches to test for three-path mediation: joint significance and bootstrapping. As stated in the main text, we used the joint significance approach. This approach has several advantages: 1) It can be easily generalized to linear mixed models (LMMs) and 2) it provides greater statistical power than the product of coefficients methods while retaining a low Type 1 error rate. Using the joint significance approach, the null hypothesis of no mediation is rejected if all three coefficients making up the indirect association are statistically significant. In our case, the three paths that make up the mediated association are: path (a) the association between perceiver's photograph-based liking and perceiver's behavioral warmth, path (b) the association between perceiver's behavioral warmth and target's behavioral warmth, and path (c) the association between target's behavioral warmth and perceiver's interaction-based liking. Since all three coefficients are required to be statistically significant, the joint significance approach provides good control of Type I error.

Despite its strengths, a limitation of the joint significance approach is it does not provide an estimate of the mediated association and thus it cannot be easily used to construct a confidence interval. The bootstrapping approach allows for constructing confidence intervals. We used the bias-corrected bootstrapping approach to construct a confidence interval around the indirect association. Specifically, we repeated the behavioral confirmation analyses estimating the indirect association based on 5,000 bootstrap resamples (Hayes, 2013) and treating the target as a fixed factor. The 95% bias-corrected confidence interval (.001, .059) excluded zero, converging with the results of the joint significance test that the three-path indirect association was statistically significant.

Supplemental Analyses: What accounts for photograph-live associations in liking judgments? Testing behavioral confirmation dynamics using self-reports

In the main text, we used independent assessments of each person's behaviors to provide evidence that behavioral confirmation accounts for the photograph-live association in liking. Although behavioral confirmation is best assessed using coding of actual behaviors, we also looked for evidence of behavioral confirmation with the self-report measures of enjoyment obtained from perceivers and targets following each interaction. Perceivers' photograph-based liking predicted their self-reported enjoyment of the interaction (*B*=.441, *SE*=.101, *p*<.001, 95% CI=[.242, .641]), which, in turn, predicted their evaluation of the target's enjoyment of the interaction (*B*=.535, *SE*=.061, *p*<.001, 95% CI=[.415, .656]), and this, in turn, predicted their interaction-based liking (*B*=.228, *SE*=.074, *p*=.003, 95% CI=[.080, .375]). This three-path indirect association accounted for 11% of the total association between photograph- and interaction-based liking. We repeated the same analyses using the target's self-reports of perceiver's enjoyment and their own enjoyment as the mediators. Perceiver's photograph-based liking of the target predicted the target's evaluation of the perceiver's enjoyment of the interaction (*B*=.222, *SE*=.088, *p*=.013, 95% CI=[.047, .398]), which, in turn, predicted the target's self-reported enjoyment of the interaction (*B*=.941, *SE*=.058, *p*<.001, 95% CI=[.827, 1.056]). However, the target's self-reported enjoyment of the interaction did not significantly predict the perceiver's interaction-based liking (*B*=-.122, *SE*=.110, *p*=.270, 95% CI=[-.341, .096]).

Supplemental Analyses: What accounts for photograph-live associations in personality judgments?

As we report in the main text, there were robust photograph-live associations in personality judgments. Perceivers who judged a target as conscientious, less neurotic, agreeable,

and more open to new experiences based on the portrait photograph judged this same target similarly following an actual live interaction. We considered possible mechanisms for such consistency, namely personality accuracy, behavioral confirmation, and a halo effect.

Personality accuracy. One possible explanation for photograph-live consistency in personality judgments is that perceivers are able to *accurately* judge personality traits based on photographs (Borkenau, Brecke, Moettig, & Paelecke, 2009). In this case, a target who is high on conscientiousness would be judged by perceivers as high on conscientiousness based on a photograph as well as based on their behaviors during a live interaction. Given that our study design completely controls for effects of targets, this account is not possible. That is, in our study, a given perceiver may have judged a target as high on conscientiousness, but a different perceiver may have judged the same target as disorganized. Because the two perceivers are judging the *same* person as possessing different personality characteristics, perceivers' judgments do not reflect the target's actual personality, but rather perceivers' idiosyncratic judgments of the target's personality.

Behavioral confirmation. Although personality is typically viewed as a relatively stable characteristic of the person, aspects of personality can be more or less salient and vary from context-to-context, as a function of the interaction partner (Zayas, Shoda, & Ayduk, 2002). Is it possible that consistency in personality judgments occurs because perceivers elicit behaviors from targets in a manner that is consistent with their initial photograph-based personality judgments (Snyder, Tanke, & Berscheid, 1977)? For example, if a perceiver judged a target as emotionally stable based on the photograph, then she might have engaged in behaviors during the live interaction to elicit greater emotional stability from the target. If so, this might have led targets to self-report different levels of a personality trait with different perceivers, and we might

be able to predict this variability as a function of perceiver's photograph-based personality judgments. But, this possibility was not supported by our data: Perceiver's photograph-based judgments of personality traits did not significantly predict targets' self-reported personality traits after the live interaction (all *t*s<1.666, all *p*s>.098).

Halo effect. Finally, we reasoned that part of the photograph-live association for personality judgments may reflect the photograph-live association for liking judgments. Based on research documenting the robustness of the halo effect, perceivers would ascribe socially desirable personality traits (Backstrom, Bjorklund, & Larsson, 2009) to targets that they like based on the photograph, and also in the live interaction, in which case personality judgments would be colored by liking judgments (Nisbett & Wilson, 1977), and hence show some consistency.

Using path analyses, we tested this possibility. Specifically, we investigated whether consistency in liking judgments explains, at least in part, consistency in personality judgments (represented in Model 1 of Fig. S2) for agreeableness, neuroticism, conscientiousness, and openness. We focused solely on these four traits because they were the ones that showed statistically significant photograph-live associations. The fit of Model 1 was compared with the alternative account that consistency in personality judgments explains the photograph-live association in liking (represented in Model 2 of Fig. S2, see Table S4 for model fit values). All models were estimated with LISREL 9.1 using the covariances among variables of interest. Model fit was evaluated using χ^2 tests. Table S4 provides all path coefficients and model χ^2 values. Evidence supporting that Model 1 is a better fit than Model 2 would be reflected by relatively smaller path coefficients and χ^2 values.

For the eight sets of path analyses tested (4 personality traits × 2 interaction types), Model 1 produced smaller χ^2 values (indicating a better fit) than Model 2 (see Table S4 for all path coefficients). Moreover, Model 1 produced χ^2 values with a $p<.10$ (indicating a poor fit) for only 2 (out of 8) models, whereas Model 2 produced χ^2 values with a $p<.10$ for 6 (out of 8) models. Finally, once interaction-based liking was partialled out, the zero-order associations between photograph-based and interaction-based personality judgments became weaker, indicating that interaction-based liking judgments at least partially mediated consistency in personality judgments. Overall, these findings support the idea that consistency in liking judgments accounts for consistency in personality judgments rather than vice versa.

Supplemental Analyses: Possible Role of Gender in Photograph-Live Association

For judgments we observed significant photograph-live associations, we repeated the analyses by including gender (0=male; 1=female) and its interaction with photograph-based judgments into the model. Although the photograph-live association in liking was evident for both males (*B*=.231, *SE*=.104, *p*=.033) and females (*B*=.616, *SE*=.080, *p*<.001), the association was stronger for females. The results regarding gender differences were mixed for personality judgments. The interaction term was significant only for neuroticism (*B*=.669, *SE*=.208, *p*=.002), with females exhibiting a stronger photograph-live association (*B*=.598, *SE*=.143, *p*<.001) than males (*B*=-.041, *SE*=.153, *p*=.792). Gender did not significantly moderate the photograph-live association in agreeableness, openness, and conscientiousness (all *p*s>.290). However, these gender differences should be treated with caution especially given that they were not observed across all outcomes and the sample size was small to detect between-gender differences.

Study S1: Lay Intuitions about the Photograph-Live Association

To assess lay intuitions about whether photograph-based judgments would predict interaction-based judgments (*beliefs about consistency vs. revision*), we recruited an Internet sample from Amazon's Mechanical Turk. One hundred and forty-nine participants accessed the survey and 114 completed it (72 female, 41 male, 1 unidentified, Mean age=38.86 years, *SD*=13.14 years). We asked participants to indicate the extent to which they believed their judgments of a target person based on a portrait photograph (i.e., head and hair only) would be indicative of how they would evaluate the target following a face-to-face interaction (1=*Strongly Disagree, 4=Neither Agree nor Disagree,* 7=*Strongly Agree*; see the Materials posted in the Open Science Framework database for the instructions and the items). Participants completed this questionnaire separately for judgments of an unknown female and an unknown male. The items were averaged separately for a female and a male target to create indices of beliefs about consistency, with higher numbers indicating that individuals expect greater consistency between photograph- and interaction-based judgments (for a female target: α=.84, *M*=2.75, *SD*=.96; for a male target: α=.85, *M*=2.88, *SD*=.99).

Results indicated that lay intuitions strongly supported a *revision hypothesis*. The indices assessing beliefs about consistency were significantly below the midpoint of the scale (for a female target: $t(113)=13.87$, $p<0.001$, $d=1.30$; for a male target: $t(113)=12.24$, $p<0.001$, $d=1.12$), suggesting that individuals strongly believed that their photograph-based judgments would be revised once they get to know a target person.

Study S2: Replicating the Photograph-Live Association with Fourteen Confederates

To provide further evidence that the photograph-live association is not due to the idiosyncratic characteristics of targets, we ran a supplemental study increasing the number of

targets from 4 to 14. Forty-nine perceivers who did not participate in the main study completed a study on first impressions for monetary compensation ($20). They were asked to interact with another participant, who in fact was one of 14 female confederates working as undergraduate research assistants (9 Caucasian, 3 Asian, 2 Half-Asian/Half-Caucasian). One perceiver who knew the target with whom they were to interact was excluded, leaving 48 perceivers (37 females).[4] Perceivers were first shown a photograph of the target (smiling in front of a neutral background) and were asked to provide impressions of her based on the photograph. They completed nine items assessing their willingness to engage in future interactions with the target (e.g., "*This seems like the kind of person whom I would like to get to know.*"; 1 = *Strongly disagree* to 7 = *Strongly agree, α*=.92) and eight items assessing favorable impressions about the confederate (e.g., "*In general, how positive is your impression of this person?*"; 1 = *Not at all* to 7 = *Very much, α*=.82; see the Materials posted in the Open Science Framework database for the items). Because these two scales were significantly correlated (r=.55, p<.001), they were averaged to index liking of the target (α=.91, M=5.07, SD=.67). Then, perceivers interacted with the target by playing a 10-minute trivia game similar to the main study. At the end of the interaction, the target left the room and perceivers completed the same items as before to report interaction-based liking of the target (α=.96, M=5.27, SD=.88). Replicating our findings, there was a significant photograph-live association in liking even when we used 14 confederates (B=.745, SE=.159, p<.001, 95% CI=[.425, 1.064], *variance explained*=31%), providing evidence that the photograph-live association is not the result of the idiosyncratic characteristics of a few confederates, and increasing the generalizability of our findings.

Footnotes

[1]Two perceiver-target pairs were given sets consisting of 15 questions due to experimenter error.

[2]The proportion of questions attempted to total questions, correct answers to total questions, or correct answers to attempted questions in the trivia game did not significantly predict liking (*t*s<1.128, *p*s>.264) or personality judgments (except for the negative association between the proportion of correct answers to total questions and openness, *B* = -2.953, *SE* = 1.310, *p*=.028; all remaining *t*s < 1.416, *p*s>.162) following the trivia game. When the proportion of questions attempted to total questions, correct answers to total questions, or correct answers to attempted questions were controlled in the analyses, the pattern of results regarding photograph-live associations in liking and personality remained the same as reported in the main text.

[3]For exploratory purposes the coders also rated how much they think the person in the video liked their interaction partner (1 = *Not at all*, 7 = *Very much*). When we included this item in the behavioral warmth composite, results of mediation analyses remained highly similar to those reported in the main text.

[4]These perceivers were preselected such that 23 indicated in an online recruitment survey that the target resembled a significant other and 25 indicated that the target did not resemble a significant other. Resemblance to the significant other did not influence photograph-based or interaction-based judgments (*t*s<1), and did not moderate the relationship between these two variables (*t*=1.06, *p*=.30) so it will not be discussed further.

References

Backstrom, M., Bjorklund, F., & Larsson, M. R. (2009). Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally. *Journal of Research in Personality, 43*(3), 335-344. doi:10.1016/j.jrp.2008.12.013

Borkenau, P., Brecke, S., Moettig, C., & Paelecke, M. (2009). Extraversion is accurately perceived after a 50-ms exposure to a face. *Journal of Research in Personality, 43*(4), 703-706. doi:10.1016/j.jrp.2009.03.007

Hayes, A. F. (2006). A primer on multilevel modeling. *Human Communication Research, 32*(4), 385-410. doi:10.1111/j.1468-2958.2006.00281.x

Hayes, A. F. (2013). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. New York: Guilford Press.

Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin, 102*(3), 390-402. doi:10.1037/0033-2909.102.3.390

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, 41*(3), 841-849. doi:10.3758/BRM.41.3.841

Letzring, T. D., Wells, S. M., & Funder, D. C. (2006). Information quantity and quality affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology, 91*(1), 111-123. doi:10.1037/0022-3514.91.1.111

Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior, 19*(3), 338-368. doi:10.1016/s0022-5371(80)90266-2

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology, 35*(4), 250-256. doi:10.1037/0022-3514.35.4.250

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. doi:10.1037/a0022885

Snijders, T. A. B., & Bosker, R. J. (1993). Standard Errors and Sample Sizes for Two-Level Research. *Journal of Educational and Behavioral Statistics, 18*(3), 237-259. doi:10.3102/10769986018003237

Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology, 35*(9), 656-666. doi:10.1037/0022-3514.35.9.656

Taylor, A. B., MacKinnon, D. P., & Tein, J.-Y. (2008). Tests of the Three-Path Mediated Effect. *Organizational Research Methods, 11*(2), 241-269. doi:10.1177/1094428107300344

Todorov, A., & Porter, J. M. (2014). Misleading First Impressions: Different for Different Facial Images of the Same Person. *Psychological Science, 25*(7), 1404-1417. doi:10.1177/0956797614532474

Zayas, V., Shoda, Y., & Ayduk, O. N. (2002). Personality in Context: An Interpersonal Systems Perspective. *Journal of Personality, 70*(6), 851-900. doi:10.1111/1467-6494.05026

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods, 12*(4), 399-413. doi:10.1037/1082-989X.12.4.399

Fig. S1. Smiling and neutral photographs of targets used in the main study. Perceivers viewed the smiling and neutral photograph of the target (separated by other questionnaires) and after viewing each photograph, they provided judgments of likability, attractiveness, and personality of the target.

Confederate 1


Confederate 2


Confederate 3


Confederate 4


Fig. S2. Hypothesized path models investigating the associations between liking and personality judgments across Time-1 and Time-2. Model 1 depicts that liking judgments shape personality judgments whereas Model 2 depicts that personality judgments shape liking judgments. Model 1 provided a better fit of the data than Model 2. For all personality traits, Model 1 produced smaller χ^2 values than Model 2. Moreover, out of the 8 sets of path analyses tested, only two had a χ^2 value with a $p<.10$ for Model 1 (one with $p<.05$, one with $p<.10$) whereas 6 (out of 8) had a $p<.10$ for Model 2 (three with $p<.10$ and thee with $p<.001$) In Model 1, after taking into account the effect of liking, the photograph-live association in personality became smaller for the trivia interaction, and it became statistically insignificant (all ps$>.05$) for the getting-to-know interaction, indicating that consistency in liking judgments at least partially accounted for consistency in personality judgments (see Table S4 for model fit values and path coefficients).

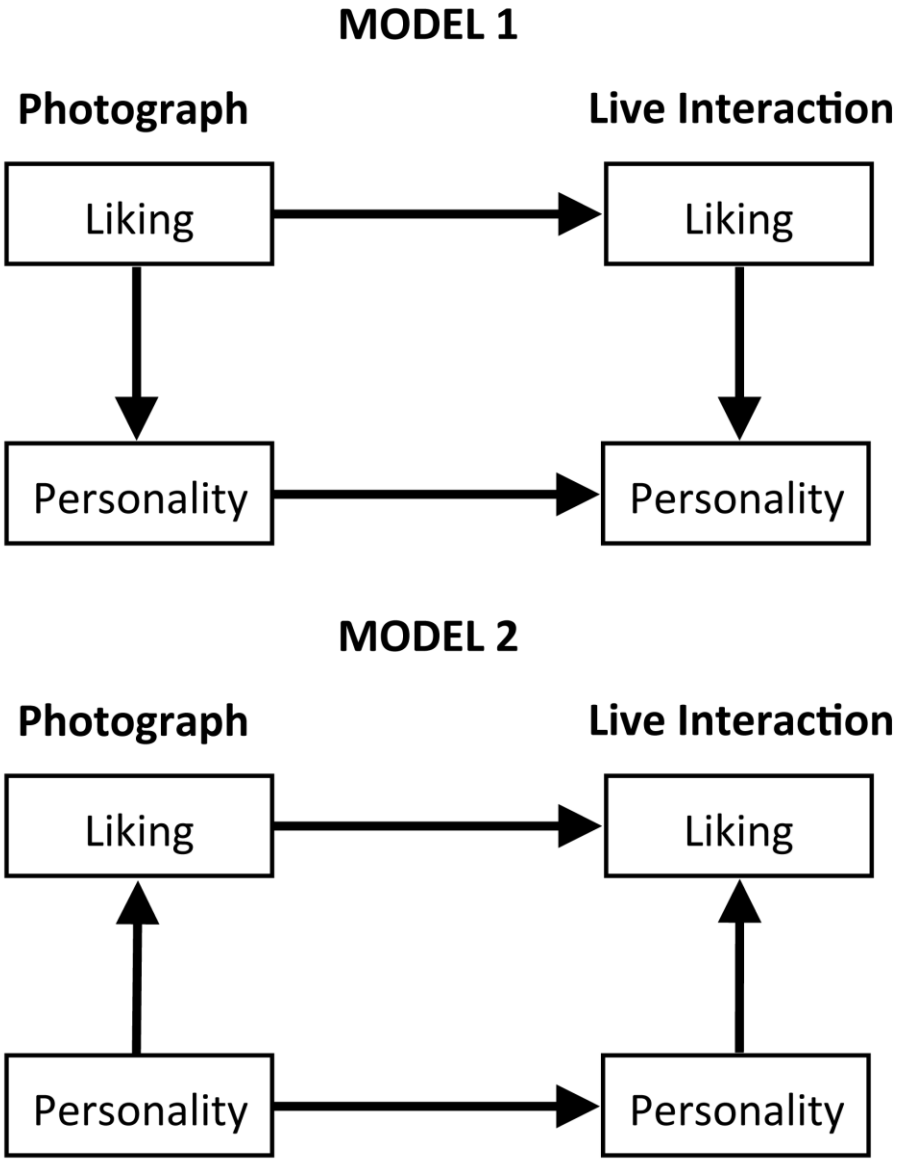

Table S1. Comparison of Photograph-Live Association across Interaction Contexts.

	95% CI (trivia minus getting-to-know)
Liking	(-0.093, 0.190)
Agreeableness	(-0.242, 0.234)
Neuroticism	(-0.115, 0.398)
Conscientiousness	**(0.025, 0.511)**
Openness	(-0.090, 0.354)

Note. The size of the photograph-live associations for trivia vs. getting-to-know judgments were compared by constructing 95% confidence intervals around the difference of the two overlapping associations obtained from the same sample (photograph-live association after the trivia interaction minus photograph-live association after the getting-to-know interaction). If the confidence interval did not include the value 0, the two associations were significantly different from each other. Consistency estimates were similar across contexts for liking, agreeableness, neuroticism, and openness. However, consistency in conscientiousness judgment was weaker after the getting-to-know interaction as compared with that after the trivia interaction. Confidence intervals indicating significant differences are in bold font.

Table S2. Perceivers' Photograph-Based Liking of Their Interaction Partner as a Function of the Valence of Their Initial Impressions (Favorable vs. Unfavorable).

	Perceivers holding favorable photograph-based impressions			Perceivers holding unfavorable photograph-based impressions		
	Number of perceivers	*M*	*SD*	*Number of perceivers*	*M*	*SD*
Target 1	7	5.69	.22	6	3.96	.11
Target 2	7	5.38	.18	7	3.24	.40
Target 3	7	5.69	.21	7	3.34	.52
Target 4	8	5.65	.25	6	3.54	.24

Table S3. Descriptive Statistics for Judgments following Live Interactions.

	Trivia		Getting-to-know	
Perceiver	*M*	*SD*	*M*	*SD*
Liking	5.51	.92	5.80	.88
Attractiveness	4.93	1.17	5.35	1.24
Extraversion	4.87	1.49	5.55	1.33
Agreeableness	5.27	1.47	5.89	1.15
Conscientiousness	4.93	1.17	5.67	1.02
Neuroticism	5.00	1.48	5.47	1.35
Openness	4.58	1.37	5.35	1.40
Own enjoyment	5.18	1.29	5.69	1.27
Partner enjoyment	5.19	1.04	5.88	.97
Target	*M*	*SD*	*M*	*SD*
Liking	5.07	.89	5.29	.98
Attractiveness	4.38	1.06	4.51	1.17
Extraversion	4.24	1.48	4.87	1.56
Agreeableness	5.15	1.25	5.65	1.06
Conscientiousness	5.38	.95	5.76	.79
Neuroticism	5.47	.88	5.71	.96
Openness	4.93	1.12	5.20	1.33
Own enjoyment	4.78	1.08	5.00	1.31
Partner enjoyment	5.03	.99	5.25	1.10

Table S4. Results of the Path Analyses Investigating the Associations between Liking and Personality Judgments

Trivia

	Model 1					***Model 2***				
	Ph L→Int L	Ph L→Ph P	Int L→Int P	Ph P→Int P	χ^2	Ph L→Int L	Ph P→Ph L	Int P→Int L	Ph P→Int P	χ^2
A.	0.62***	0.68***	0.40***	0.28*	3.50	0.52***	0.68***	0.33**	0.42***	4.92†
N.	0.62***	0.63***	0.45***	0.23†	0.35	0.50***	0.63***	0.34***	0.39**	4.75†
C.	0.62***	0.50***	0.43***	0.36***	1.10	0.48***	0.50***	0.37***	0.53***	3.13
O.	0.62***	0.21	0.41***	0.35***	0.51	0.56***	0.21	0.30**	0.39**	5.00†

Getting-to-Know

	Model 1					***Model 2***				
	Ph L→Int L	Ph L→Ph P	Int L→Int P	Ph P→Int P	χ^2	Ph L→Int L	Ph P→Ph L	Int P→Int L	Ph P→Int P	χ^2
A.	0.58***	0.68***	0.63***	0.17	0.68	0.35***	0.68***	0.56***	0.42***	3.50
N.	0.58***	0.63***	0.71***	0.04	4.78†	0.27**	0.63***	0.62***	0.25†	18.59***
C.	0.58***	0.50***	0.48***	0.08	7.70*	0.45***	0.50***	0.29**	0.26*	16.41***
O.	0.58***	0.21	0.66***	0.18†	0.30	0.37***	0.21	0.56***	0.25†	12.51***

Note. A=Agreeableness; N=Neuroticism; C=Conscientiousness; O=Openness; Ph L→Int L=Photograph-based liking judgment predicting interaction-based liking judgment; Ph L→Ph P=Photograph-based liking judgment predicting photograph-based personality judgment; Int L→Int P=Interaction-based liking judgment predicting interaction-based personality judgment; Ph P→Int P=Photograph-based personality judgment predicting interaction-based personality judgment; Ph P→Ph L=Photograph-based personality judgment predicting photograph-based liking judgment; Int P→Int L=Interaction-based personality judgment predicting interaction-based liking judgment. All path coefficients are standardized. Model fits were evaluated using a χ^2 test at 2 degrees of freedom. Lower χ^2 values indicate better fit. For all personality traits, Model 1 produced smaller χ^2 values than Model 2. Moreover, out of the 8 sets of path analyses tested, only two had a χ^2 value with a $p<.10$ for Model 1 whereas 6 (out of 8) had a $p<.10$ for Model 2. Overall, these findings indicate that Model 1 provides better fit to the data than Model 2. That is, the data support the hypothesis that consistency in liking judgments accounts for consistency in personality judgments rather than vice versa. Once consistency in liking is taken into account, the zero order associations between photograph-based and interaction-based personality judgments (see Ph P→Int P on Column 10) get weaker (as shown under Ph P→Int P on Column 5).
*** $p<.001$, ** $p<.01$, * $p<.05$, † $p<.10$.

Acknowledgments

The authors would like to thank Steve Strycharz, Gizem Surenkok, Jeffrey McCalla, and Rachel Cooper for their feedback on earlier drafts, Katrina Simon, Jeffrey McCalla, Samantha Chanko, Christina Haywood, Stephanie Izard, Aaron Lee, Brooke Neider, and Emily Schreiner for their assistance with coding the videos, and Cornell Statistical Consulting Unit for their assistance with data analyses.