Empirical Research Paper

Everyday Perceiver-Context Influences on Impression Formation: No Evidence of Consistent Effects

Sally Y. Xie[1] (iD)**, Sabrina Thai[2]** (iD)**, and Eric Hehman[1]**

Abstract

Facial impressions (e.g., trustworthy, intelligent) vary considerably across different perceivers and targets. However, nearly all existing research comes from participants evaluating faces on a computer screen in a lab or office environment. We explored whether social perceptions could additionally be influenced by perceivers' experiential factors that vary in daily life: mood, environment, physiological state, and psychological situations. To that end, we tracked daily changes in participants' experienced contexts during impression formation using experience sampling. We found limited evidence that perceivers' contexts are an important factor in impressions. Perceiver context alone does not systematically influence trait impressions in a consistent manner—suggesting that perceiver and target idiosyncrasies are the most powerful drivers of social impressions. Overall, results suggest that perceivers' experienced contexts may play only a small role in impressions formed from faces.

People form impressions of one another at a glance, such as whether a person looks high-status or intelligent (Bjornsdottir & Rule, 2017; Zebrowitz et al., 2003). Regardless of accuracy, these impressions are pervasive and consequential, predicting election outcomes (Ballew & Todorov, 2007; Olivola & Todorov, 2010), sentencing decisions (Blair et al., 2004; Wilson & Rule, 2015), and financial lending rates (Duarte et al., 2012) in the real world. Yet our understanding of the way that facial impressions are formed depend overwhelmingly on face ratings made by people situated in social psychology studies. In most of these studies, people rate faces while sitting in front of a computer, in a highly controlled laboratory environment. Whether impression formation differs when participants experience other contexts is unknown—even though this question is critical to how we should interpret such lab-based results. To what extent are facial impressions influenced by situational and contextual factors that participants experience in daily life? We adopted an experience-sampling paradigm (Thai & Page-Gould, 2018), collecting ratings from participants as they went about their day, to address this question.

Variability in Facial First Impressions

Social impressions from faces are jointly influenced by perceiver characteristics, target characteristics, and perceiver-by-target interactions (Hehman et al., 2018; Kenny, 2019;

Kunda et al., 1996). Among these components, the contributions of target characteristics (e.g., morphological cues, social identity) have been studied extensively in isolation, with hundreds of studies demonstrating how different targets elicit judgments of attractiveness, trustworthiness, among many other traits (Todorov et al., 2015). This literature is situated in ecological theories that highlight the functional significance of face perception, offering a partial explanation for the human tendency to readily overgeneralize facial cues (e.g., an upturned mouth) to stable trait inferences (e.g., friendly; Zebrowitz et al., 2003). More recent work has also examined how targets in different contexts are perceived, by varying the visual context of target stimuli. For example, people integrate facial cues (e.g., untrustworthy face) and contextual cues (e.g., threatening or neutral scene) when evaluating the trustworthiness of a face (Brambilla et al., 2018; Mattavelli et al., 2021), and faces appear more attractive when they appear in a group (Carragher et al., 2021).

Yet perceivers also play an active role in impression formation, differing in their impressions of the same face. For

[1]McGill University, Montreal, Quebec, Canada
[2]Brock University, St. Catharines, Ontario, Canada

Corresponding Author:
Sally Y. Xie, Department of Psychology, McGill University, 2001 McGill College Ave., Montreal, Quebec, Canada H3A 1G1.
Email: sally.xie@mail.mcgill.ca

instance, perceivers who vary in their social identity (Kawakami et al., 2017) or stereotype knowledge (Oh et al., 2019; Wilson et al., 2017) may evaluate the same target very differently. These perceiver contributions are central to modern theories of social cognition (Brewer, 1988; Macrae & Bodenhausen, 2000), and current perspectives conceptualize impression formation as a dynamic process, during which the bottom-up processing of facial features interacts with multiple top-down cognitive factors (Freeman et al., 2020). For example, people's intuitions about trait correlations (e.g., "how intelligent is someone who is attractive?") explain considerable variability in how facial impressions are formed (Stolier et al., 2018, 2020), suggesting that conceptual knowledge unique to each perceiver shapes impression formation.

Recently, researchers have examined the relative importance of these components in face impressions (Hehman et al., 2017; Hönekopp, 2006; Judd et al., 2012; Xie et al., 2019). By using cross-classified multilevel models to estimate variance components from different clusters in the data, this research decomposes the total variance in face impressions to those uniquely attributable to perceivers, targets, or perceiver-by-target interactions (Hehman et al., 2017; Judd et al., 2012; Kenny, 2019; Xie et al., 2019). Results indicate perceiver idiosyncrasies contribute a greater proportion (~20%–25%) of variance than target characteristics (~10-15%), though perceiver-by-target interactions contribute the most overall (~32-39%; Hehman et al., 2017; Hönekopp, 2006; Xie et al., 2019).

Although perceiver characteristics appear to contribute a large share of known variance in face impressions, it is unclear what this "perceiver-level variance" captures. It may reflect stable, trait-level idiosyncrasies such as personality, stereotype knowledge, response style, or state-like influences such as affective state, evaluative context, external environment, or psychological situations. Recent work suggests that perceivers' stereotype knowledge and lay theories of personality play a role (Stolier et al., 2020, 2018; Xie et al., 2021), as well as their degree of acquiescence and positivity bias (Heynicke et al., 2021). However, these factors do not explain all perceiver variance, and other sources of perceiver-level variance are likely important. Critically, between 20% and 40% of the variance in facial impressions remains unexplained. This unexplained variance may reflect measurement error, or a meaningful source of intraindividual variability that has yet to be explored. To that end, the present research examines to what extent situational, day-to-day contextual factors experienced by perceivers contribute to variability in social impressions.

Do Perceiver Contexts Influence Impression Formation?

Impression formation does not occur in a vacuum. In everyday life, people are embedded in various contexts when forming impressions of others. A perceiver's context can encompass one's broader culture (Jaeger et al., 2019), personal environment (Barrett & Kensinger, 2010), or experienced situation (Rauthmann & Sherman, 2018). Although research on the influence of perceiver contextual factors on impression formation is scarce, a recent twin study found that genes explain little variability in facial impressions compared to one's personal environment (Sutherland et al., 2020), encompassing local factors related to one's upbringing and community. Consistent with this finding, research with a large, international sample found that the broader cultural context also explains minimal variability, relative to individual differences (Hester et al., 2021). These findings allow for the possibility that any meaningful perceiver-level contextual variability in face impressions may exist at the locus of situational, day-to-day variation in one's recent experiences, rather than in one's broader culture or genetic makeup.

To the extent that these everyday experiential factors are psychologically meaningful, they may impact the impression formation process. For example, in contexts associated with harm (e.g., weapons are present), people readily evaluate others as angrier (Holbrook et al., 2014; Maner et al., 2005), larger (Fessler et al., 2012), and more physically threatening (Wilson et al., 2017), compared to neutral contexts. Furthermore, perceivers' mood states may interact with features of the environment to impact situation construal. For example, people form impressions that are mood-congruent (Abele & Petzold, 1994; Forgas, 1992; Forgas & Bower, 1987), and properties of the environment can both shape and be shaped by mood (Chartrand et al., 2006). The psychological experience of perceivers may therefore impact the way that they process and interpret novel targets.

To experimentally assess how impression formation varies across a sufficiently diverse number of perceiver contexts would require there to be an improbably large number of fixed situations, in which we manipulate participants' moods, perceived situations, and environment. Instead of experimentally inducing these different contexts, we used an experience-sampling paradigm (Thai & Page-Gould, 2018) to explore how perceivers—going about their daily lives and experiencing different contexts in a naturalistic manner—form impressions of targets.

To our knowledge, no studies have examined the influence of daily experiences on perceivers forming impressions, and there are no theoretical frameworks from which to derive specific hypotheses. With notable exceptions, most of the existing research on impression formation comes from participants embedded in the context of a social psychology experiment, sitting in a lab and rating faces. Furthermore, participants typically only evaluate each target once, which limits the amount of intraindividual variability that can be observed. Some research has examined context in impression formation, but focusing on target contexts, with targets embedded in diverse visual contexts as impressions are formed (e.g., Brambilla et al., 2018; Carragher et al., 2021;

Fessler et al., 2012). Accordingly, here we present the first exploratory study to examine the perceiver context factors that might influence how they form impressions.

As participants went about their daily lives, we sent them a photo of a face, and measured their impressions as well as aspects of their physical and psychological context. We aimed to answer two research questions: First, to what extent do perceivers' everyday contexts matter for impression formation? Second, which perceiver contexts are important for driving impressions in a systematic manner across different participants? By collecting evaluations from participants over time, we allowed natural sources of intraindividual variability to emerge, and examined whether perceivers' contexts meaningfully contributed to variability in social impressions.

Methods

Experience Sampling

We explored the impact of people's naturally varying, day-to-day contexts on the way that they form impressions from faces, using experience sampling (Thai & Page-Gould, 2018) to track daily changes in participants' contexts at the moment that they form impressions of facial stimuli. We used "perceiver context" broadly to encompass perceivers' environmental and psychological states that might contribute to intraindividual variability in facial impressions. Accordingly, we focused on state-like variables that were likely to fluctuate within individuals. Data, code, and study materials available at [osf.io/xdmjr]. This research was approved by the McGill University Research Ethics Board.

Participants

330 U.S. participants were recruited from Amazon Mechanical Turk to complete an intake questionnaire and participate in the experience-sampling study. We overrecruited to ensure we would be able to attain a final sample similar to previous research that used this experience-sampling method (Thai & Page-Gould, 2018). In all, 218 participants (52% female, M_{age} = 36.0, SD_{age} = 10.3) continued with the study: 168 White, 18 East Asian, 17 Black, 8 Latine, 3 Aboriginal/Indigenous, 2 multiracial, 1 South Asian, and 1 undisclosed. The average income of participants was $61,148 ($SD$ = $35,269), and the highest level of education attained included 1 high school or less, 19 high school graduate, 52 some college, 33 associate's degree, 77 bachelor's degree, 31 master's degrees, 1 professional degree, 3 doctoral degree, and 1 undisclosed.

Procedure

Participants completed an intake questionnaire, which included a brief measure of personality (Rammstedt & John, 2007) and demographic questions. They were then directed to a webpage explaining how to install and use the *ExperienceSampler* smartphone app (Thai & Page-Gould, 2018) that would record their responses throughout the day for up to 15 consecutive days. Figure 1 reveals that participants completed measures across a wide variety of geographic locations, ensuring variability in some of the factors that we measured, such as weather or other characteristics of the environment.

Data collection proceeded in multiple waves between August 29 and December 16 in 2019. Participants were notified twice a day for up to 15 consecutive days at quasi-random times via the app. Participants indicated the hours that they would be available to use their phone on weekdays and weekends, and notification times were randomized within these periods. When responding to a notification, participants were asked to report their impression of one randomly-selected human face, on 6 traits commonly assessed in impression formation research: friendliness, trustworthiness, attractiveness, intelligence, physical strength, and dominance (Hehman et al., 2017; Todorov et al., 2015). Participants rated each trait impression on 1-"Not at all" to 7-"Very much" Likert-type scales. Traits were presented in randomized order across surveys and across participants. After ratings, participants then completed brief measures of their current situation, environment, mood, and physiological state. The order of questionnaires and items within questionnaires were also randomized per survey. Measures were brief to minimize participant fatigue and attrition. See Supplementary Materials for the complete questionnaire.

Our theoretical and statistical focus was on within-subject variability, and a limited set of stimuli was ideal for focusing on the potential influence of perceiver contexts. See Figure 2 for an illustration of within-subject variability in target ratings over time. Stimuli comprised photographs of White, emotionally-neutral faces (8 male, 7 female) randomly selected from the Chicago Face Database (Ma et al., 2015). If participants responded to all notifications, then a total of 180 trait ratings would be obtained per participant: 30 responses × 6 trait ratings of 15 stimuli, each rated twice. Participants could miss 5 notifications total, after which they would no longer receive notifications.

Measures. We measured several ways in which physical environment and psychological states may vary within individuals across 15 days. However, there are countless factors that might influence one's immediate psychological context. Because it was not possible to comprehensively explore all contexts, items were selected based on salience and subjective intuitions by the research team, with the goal of casting a wide net. In total, measures of perceivers' experienced contexts included 22 items, administered at each survey. No variables other than those reported in the manuscript have been measured. See Supplementary Materials for the full questionnaire.

Measuring situations. Modern approaches to situation measurement focus on how situations are subjectively
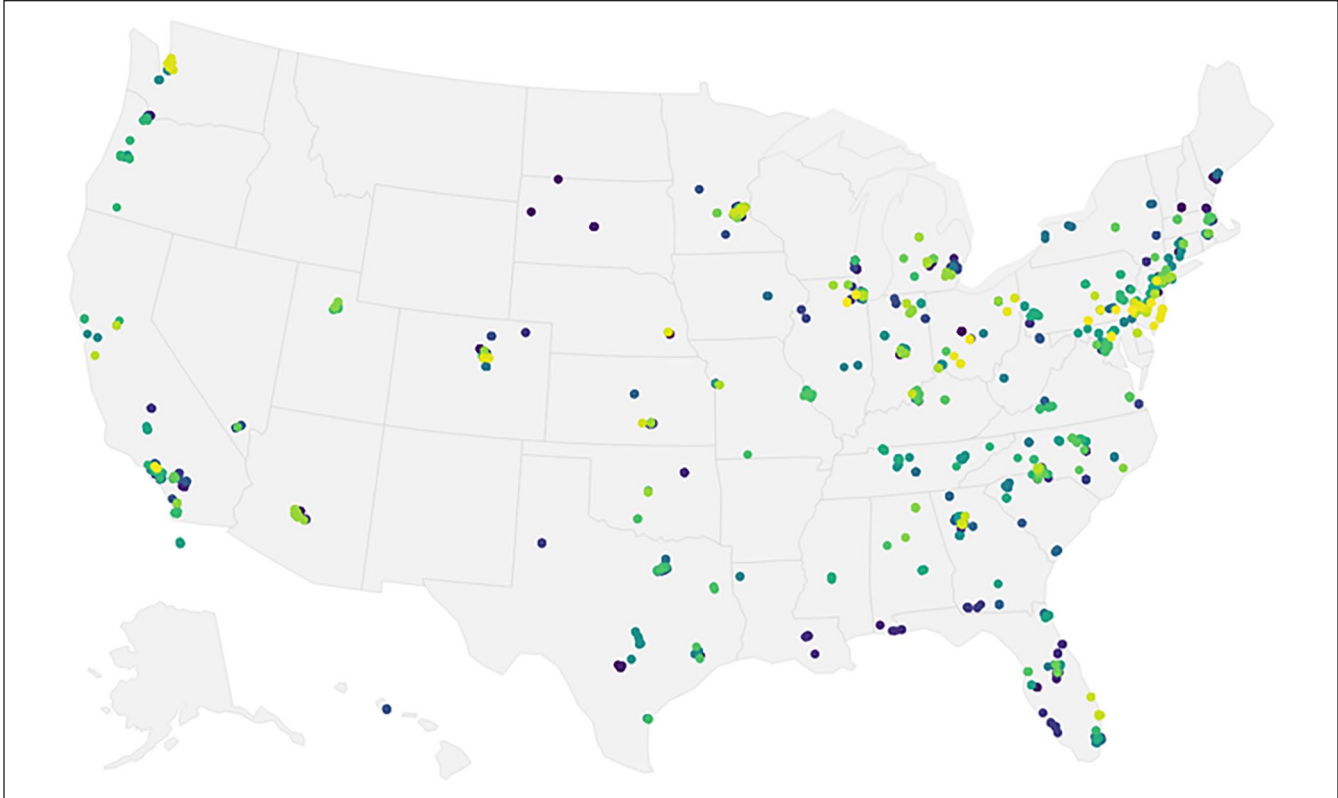
Figure 1. Participants' location (longitude/latitude) at the time of each response (constrained to the U.S. for visualization).
Note. Unique colors represent unique participants (*n* = 218).
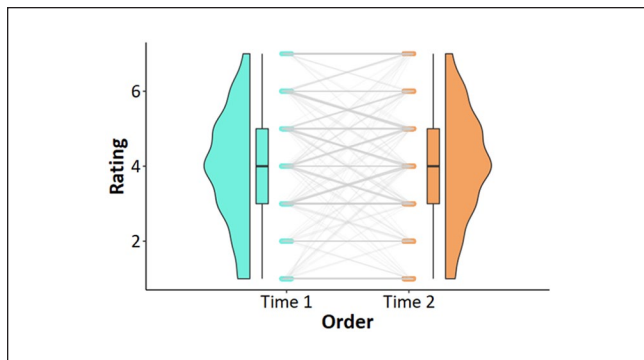
Figure 2. Ratings of the same target at Time 1 and Time 2, across all participants and traits.
Note. Gray lines connect ratings from the same participants. That many are diagonal reveals that participants did not always provide identical ratings to the same target across different time points, allowing for the possibility that perceivers' experienced contexts might explain such variation.

perceived (Brown et al., 2015; Parrigon et al., 2017; Rauthmann & Sherman, 2018). Since 2014, various situation taxonomies have been developed around this principle. Given the large content overlap across different measures (Horstmann et al., 2017), we decided to use the shortest validated measure available for the systematic assessment

of situations that focuses specifically on the description of everyday situations: the ultra-brief (8-item) form of the Situational 8 DIAMONDS (Rauthmann et al., 2014; Rauthmann & Sherman, 2015), a taxonomy of situation characteristics comprising (D)uty, (I)ntellect, (A)dversity, (M)ating, p(O)sitivty, (N)egativity, (D)eception, and (S)ociality. The ultra-brief form of the DIAMONDS (Rauthmann & Sherman, 2015) has one item tapping each dimension (e.g., "Are you in a situation where work has to be done?"), on 1-"Not at all" to 7-"Totally" Likert-type scales.

Measuring environment. Environment variables included weather (i.e., sunny, rainy) measured on 1-"Not at all" to 7-"Very much" Likert-type scales, temperature (in Fahrenheit, on a sliding scale from Very Cold: -20 to Very Hot: 120), and checkboxes indicating whether the respondent was indoors or outdoors, alone, with strangers, or with familiar others.

Measuring mood. We included six items capturing mood: happy, calm, energetic, fearful/anxious, angry, and sad. These were derived from adjectives loading strongly on the mood factors identified in the UWIST Mood Adjective Checklist (Matthews et al., 1990). Participants responded to the prompt, "Thinking about yourself and how you feel in

the past 30 minutes, to what extent do you feel: [. . .]," on a Likert-type scale from 1-"Not at all" to 7-"Very much."

Measuring physiological state. We included two items capturing basic physiological states: Tired and Hungry. Participants responded to the prompt, "How [tired / hungry] are you right now?" on a 1-"Not at all" to 7-"Very much" Likert-type scale.

Demographics. In the intake questionnaire, participants completed a variety of demographic items (i.e., gender, age, ethnicity, income, and education). In addition, they completed a brief measure of personality: the 11-item version of the Big Five Inventory (Rammstedt & John, 2007). Given our interest in within-subject effects, these variables were not of primary interest, but served as robustness checks to ensure that any contextual effects were not explained by trait-like perceiver characteristics.

Analysis 1

Quantifying Perceiver-Context Variability in Face Impressions

We quantified the variance in facial impressions attributable to different contexts experienced by perceivers. Because we had no a priori hypotheses about which perceiver contexts would influence facial impressions, nor to what extent they could influence impressions, we randomly (by participant) partitioned the data into exploratory ($N = 109$, n ratings = 2236) and confirmatory hold-out ($N = 109$, n ratings = 2190) datasets to reduce the possibility that any models developed on the exploratory data were overfitted. There were no preregistrations for this study.

Analytic approach. We built a cross-classified multilevel model with no predictors (i.e., null model) to partition the data into variance attributable to context, perceiver, target, and their higher order interactions. Similar cross-classified models have been used in social psychology research (Judd et al., 2012) to decompose and quantify the variance in impressions originating at the target and perceiver levels (Hehman et al., 2017; Hönekopp, 2006; Kenny, 2019; Xie et al., 2019). Here, the level-1 unit of analysis is a trait rating made at the time that participants are responding to each survey, which is cross-classified by targets, perceivers, and contexts. Models were estimated using the *lme4* (Bates et al., 2015) package in R.[1] See Supplementary Materials for further elaboration of this model.

Modeling heterogeneity in real-world contexts. To estimate this model, each rating in Level 1 of the model must be nested within a categorical context cluster, just as it is nested within a participant and target cluster. Given that there are 22 predictors, it is impractical to model all higher-order interactions by

entering them as predictors in a multilevel model. For example, including all higher-order interactions requires estimating an additional 4,194,281 parameters. Thus, an important first step was to identify distinct perceiver contexts, and assign each response to a distinct context in a class of contexts. For example, one perceiver context (e.g., outdoors, warm, sunny, social environment, hungry) may be differentiated from another perceiver context (e.g., alone, indoors, in an environment that requires work to be done, tired) based on participants' responses to multiple contextual variables. As we did not have any a priori hypotheses about which combinations of perceiver-level contextual variables might be psychologically meaningful, we adopted a data-driven approach.

Our strategy was to identify qualitatively distinct perceiver contexts that emerge from combinations of contextual features. Specifically, we used latent profile analysis (LPA) to examine how participants' responses to these contextual variables cluster together and constructed distinct classes of contexts in a data-driven manner, using quantitative data to express qualitatively distinct contexts. This was possible given a longitudinal dataset, with observations that are repeated within (and between) participants who differ in trait characteristics (e.g., personality, worldview) but who may nonetheless experience similar psychological states as they experience similar contexts. This strategy allowed us to estimate the variance in social impressions arising from perceiver contexts. We implemented LPA using the *tidyLPA* package in R (Rosenberg et al., 2019).

Latent Profile Analysis of Real-World Contexts

LPA estimates an underlying categorical latent variable from continuous indicators (Hox & Roberts, 2011; Pastor et al., 2007). Often used in person-centered analyses, its practical advantage is that it mimics higher-order interaction terms and catalogs complicated interaction effects in a simple way, as subgroups or "classes." This approach is suited for data in which distinct subgroups—that is, qualitative differences—are expected (Hox & Roberts, 2011; Pastor et al., 2007). Here, the 22 contextual indicators intend to describe qualitatively distinct real-world contexts experienced by participants when they respond to each survey. LPA has previously been used to examine distinct subtypes in personality (Merz & Roesch, 2011) and goal orientation (Pastor et al., 2007).

Some common concerns about these class of models include the sensitivity of the class separation and the number of latent profiles correctly identified (Bauer & Curran, 2003; Peugh & Fan, 2013). To maximize the correct identification of latent profiles, we used a class-invariant unrestricted parametrization, which offers some improvement in model recovery over the default of assuming local independence (Pastor et al., 2007; Peugh & Fan, 2013). In determining sample size, we ensured that the number of observations would greatly exceed $n = 500$ even after partitioning into an exploratory ($n = 2236$) and confirmatory segment ($n = 2190$), given the within-subjects

design. Next, we searched between 2 and 51 classes to cover a broad range of possible classes (51 is a computational ceiling). Model selection was based on two indices: the Bootstrap Likelihood Ratio Test (BLRT) and the Bayesian Information Criterion (BIC), determined in a recent simulation study to outperform other indices in terms of correctly and reliably recovering the true number of classes across different sample sizes (Nylund et al., 2007). The BIC balances goodness-of-fit with parsimony (Raftery, 1995); reductions of 10 points or more between two models indicates improved fit. The BLRT compares the fit between two models, where *p*-values below .05 indicate superior fit of class *k* versus *k* – 1. Finally, we made decisions on number of latent profiles based on both the exploratory and confirmatory dataset.

LPA assigns each survey response to a certain class based on the highest probability of belonging to each class. We used the exploratory data to search for the optimal number of classes based on the lowest BIC and significance on the BLRT. Because LPA was designed to model heterogeneity in observed data, it was unlikely that the optimal number of classes should replicate exactly across data with different inputs. However, we expected the optimal number of classes to be similar across exploratory and confirmatory segments of our data. By identifying the best-performing model in the exploratory dataset and validating its performance in the confirmatory dataset, we could be more confident that LPA had retrieved the correct number of latent profiles (i.e., perceivers' experienced contexts in which impressions are formed) from the observed variables.

In the main analysis, we entered this perceiver-contextual class variable into the cross-classified model as a random cluster, along with perceivers and targets. Estimates from these models were used to calculate intraclass correlation coefficients (ICCs). These ICCs represent the percentage of variance in a trait rating explained by different clusters of the multilevel model.

Results

We implemented LPA to identify the number of qualitatively distinct perceiver-contexts observed in our data. LPA conducted on the exploratory dataset (*n* ratings = 2,236, *N* participants = 109) found the optimal number of contexts to be 44. We assessed the robustness of this solution with the confirmatory dataset (*n* ratings = 2,190, *N* participants = 109), which found the optimal number of contexts to be 50, followed by 42 and 44. See Supplementary Table 1 for the top 10 solutions ranked by lowest BIC across both datasets.

Of the solutions that had a significant BLRT *p*-value across both datasets and the lowest BIC values, the model with 44 classes was the most parsimonious. We therefore selected the 44-class model for our primary analyses. We do not interpret these 44 classes as a representative, generalizable taxonomy of real-world contexts experienced by perceivers, but rather the number of distinct perceiver contexts

present in our dataset. This allowed us to include the contextual cluster (i.e., with 44 distinct contexts) as a random cluster in a multilevel model. We conducted supplementary analyses with other class solutions which had significant *p*-values on the BLRT to confirm that results weren't contingent on a particular solution (Supplementary Analysis 1).

In our main analysis, each survey response was assigned to a "context" class based on the LPA solution with 44 classes. On average, participants in the exploratory dataset experienced 8.83 total contexts (*SD* = 4.01, *range* = 1–18), whereas participants in the confirmatory dataset experienced 8.59 total contexts (*SD* = 3.44, *range* = 1–17). See Supplementary Table 3 for an example of one distinct "context" that was identified according to this classification.

Quantifying contextual variability in face impressions

Overview. The variance in trait impressions across all 6 traits was decomposed into between-perceiver, between-target, between-context, perceiver × target, perceiver × context, target × context, and residual variance. We first present a bird's-eye view of all ICC estimates and 95% CIs (Table 1) from the exploratory (Figure 3) and confirmatory (Figure 4) datasets.

Contributions of perceivers' experienced contexts. Central to our research question, the novel aspect of this analysis relates to the unique contribution of day-to-day perceiver contexts to face impressions. We found that the contextual factors examined here do not, on their own, contribute any unique variance to face impressions (~0%). This indicates that the average rating made in one context class (across all perceivers rating all targets) does not differ from the average rating made in another context class (across all perceivers rating all targets). For example, in a simplified scenario in which being in a sunny setting or not was a distinct context experienced by perceivers, if ratings were consistently different when perceivers were in a sunny versus less sunny setting, then we would observe a higher context-ICC.

Importantly, perceivers' experienced contexts do not meaningfully contribute variance to trait impressions regardless of the perceiver or target being rated. Summarizing across all traits, the perceiver × context interaction ICC contributed only ~1.6% (exploratory: 0.7%—3.2%) and ~0.5% (confirmatory: 0.0%—1.1%) of the variance in face impressions. This suggests that different participants experiencing different contexts did *not* vary in their trait ratings (regardless of which stimuli they were evaluating). As a hypothetical example, if happy people evaluated others as friendlier on a sunny day, whereas unhappy people evaluated others as less friendly on such a day, then we would observe a higher perceiver-by-context ICC. In this scenario, differences between perceivers (how happy they are on average) interact with their experienced contexts (how sunny it is when they respond to the survey) to shape their impressions of any target's face. However, these perceiver × context interactions

Table 1. Intra-Class Correlation Coefficients and 95% Confidence Intervals by Trait.

Trait	Exploratory dataset (*n* = 109)				Confirmatory dataset (*n* = 109)			
	Perceiver ICC	95% CI	Target ICC	95% CI	Perceiver ICC	95% CI	Target ICC	95% CI
Trustworthy	.231	[.163, .293]	.162	[.052, .255]	.235	[.169, .297]	.151	[.051, .238]
Friendly	.201	[.133, .262]	.223	[.096, .349]	.182	[.120, .236]	.235	[.105, .360]
Attractive	.238	[.169, .305]	.153	[.047, .243]	.252	[.183, .319]	.142	[.040, .224]
Intelligent	.241	[.177, .304]	.091	[.018, .147]	.213	[.155, .271]	.105	[.026, .170]
Dominant	.124	[.075, .168]	.224	[.094, .347]	.136	[.085, .179]	.167	[.056, .265]
Physically strong	.163	[.106, .215]	.209	[.084, .327]	.183	[.124, .238]	.180	[.063, .283]

Trait	Exploratory dataset (*n* = 109)				Confirmatory dataset (*n* = 109)			
	Perceiver × target ICC	95% CI	Context ICC	95% CI	Perceiver × target ICC	95% CI	Context ICC	95% CI
Trustworthy	.238	[.188, .287]	.001	[.000, .003]	.198	[.148, .244]	.000	[.000, .000]
Friendly	.236	[.178, .287]	.002	[.000, .004]	.228	[.169, .280]	.000	[.000, .000]
Attractive	.363	[.301, .423]	.000	[.000, .000]	.280	[.225, .330]	.000	[.000, .000]
Intelligent	.269	[.218, .316]	.005	[.000, .010]	.202	[.152, .250]	.000	[.000, .000]
Dominant	.324	[.256, .388]	.001	[.000, .003]	.291	[.231, .348]	.000	[.000, .000]
Physically Strong	.285	[.221, .345]	.003	[.000, .005]	.244	[.188, .293]	.000	[.000, .000]

Trait	Exploratory dataset (*n* = 109)				Confirmatory dataset (*n* = 109)			
	Perceiver × context ICC	95% CI	Target × context ICC	95% CI	Perceiver × context ICC	95% CI	Target × context ICC	95% CI
Trustworthy	.020	[.000, .040]	.011	[.000, .021]	.000	[.000, .000]	.010	[.000, .020]
Friendly	.032	[.007, .056]	.003	[.000, .005]	.000	[.000, .000]	.000	[.000, .000]
Attractive	.015	[.000, .030]	.002	[.000, .005]	.002	[.000, .004]	.008	[.000, .016]
Intelligent	.007	[.000, .013]	.000	[.000, .000]	.009	[.000, .018]	.013	[.000, .027]
Dominant	.015	[.000, .030]	.012	[.000, .024]	.011	[.000, .021]	.009	[.000, .017]
Physically Strong	.008	[.000, .016]	.000	[.000, .000]	.008	[.000, .017]	.018	[.000, .036]

Note. ICC = intraclass correlation coefficients; CI = confidence interval.
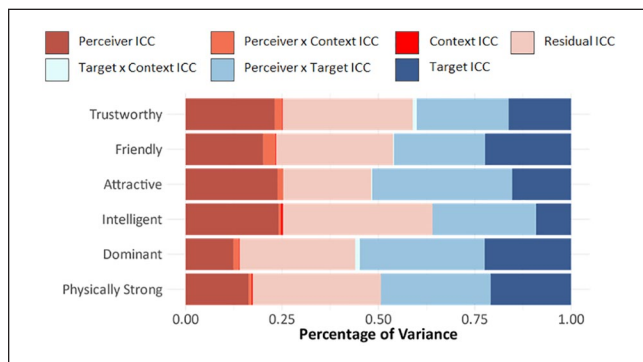
Figure 3. Relative contributions of perceiver-level, target-level, context-level, perceiver × target, perceiver × context, target × context, and residual variance to trait impressions: trustworthy, friendly, attractive, intelligent, dominant, and physically strong. *Note.* Results from exploratory dataset. ICC = intraclass correlation coefficients.
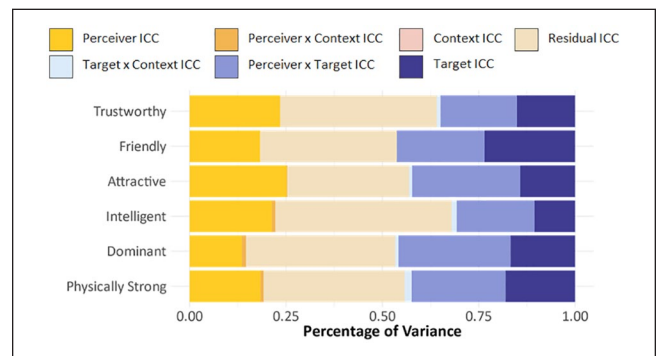
Figure 4. Relative contributions of perceiver-level, target-level, context-level, perceiver × target, perceiver × context, target × context, and residual variance to trait impressions: trustworthy, friendly, attractive, intelligent, dominant, and physically strong. *Note.* Results from confirmatory dataset. ICC = intraclass correlation coefficients.

contributed very little variation in trait impressions, indicating that different participants were not differentially affected by their day-to-day contexts when forming impressions of strangers. Similarly, the target × context ICC contributed only ~0.5% (exploratory: 0.0%—1.2%) and ~1.0% (confirmatory: 0.0%—1.8%) of variance, suggesting that different

targets being rated in different perceiver-contexts did *not* elicit different trait ratings (regardless of rater). As a hypothetical example, if targets with downturned eyebrows were evaluated as more intelligent when raters were in a work situation, whereas targets with upturned eyebrows were rated as less intelligent in such a situation, then we would observe a higher target-by-context ICC. However, differences between targets (e.g., eyebrow shape) do not appear to interact with any perceiver's experienced context (being in a work situation) to shape impressions of the target. We discuss the implications of these findings in more detail in the General Discussion.

Perceiver versus target contributions. Other contributions were generally consistent with previous research (Hehman et al., 2017; Hönekopp, 2006; Xie et al., 2019). Summarizing across 6 traits, results indicated that between-perceiver differences uniquely contributed ~20% of the variance in face impressions in the exploratory dataset (12.4%–23.8%) and ~20% in the confirmatory dataset (13.6%–25.2%). These contributions varied across traits in a manner consistent with previous literature. Between-target differences (e.g., facial appearance) uniquely contributed, on average, ~17.7% (exploratory: 9.1%–22.4%) and ~16.4% (confirmatory: 10.5%–23.5%) of the variance in each dataset. Both the perceiver-ICC and target-ICC estimates generally replicated previous work partitioning variance in face impressions (Hehman et al., 2017; Xie et al., 2019).

Across all sources of variance, the perceiver × target ICC was the largest in both exploratory and confirmatory segments. Summarizing across 6 traits, this interaction contributed ~29.0% (exploratory: 23.6%–36.3%) and ~24.0% (confirmatory: 19.8%–29.1%) of the variance in face impressions. This estimate was similar but slightly smaller than the estimates observed in previous work examining the perceiver × target interaction component (Hehman et al., 2017; Hönekopp, 2006). This perceiver × target interaction can be interpreted as "personal taste," or differential criteria that perceivers use when judging different stimuli.

Robustness checks. Given our design, one additional concern was that rating faces twice over the 15-day period might have influenced results. Exploring this possibility, estimates did not change when we additionally included a variance component for participation-over-time, using participants' chronological trial count (see Supplementary Analysis 1C). This suggests that participation in the study over time did *not* introduce any variability in responding (e.g., as a result of fatigue or boredom). We also checked whether repeated presentations of a target (i.e., the exact same photo) influenced subsequent ratings of that same target. Overall, naïve ratings do not differ significantly from subsequent ratings in a systematic manner when averaging across perceivers, targets, and contexts (Supplementary Analysis 1D). This indicates that the mere act of seeing the same face again did

not systematically shift trait ratings. Finally, we added to the model the number of unique contexts that each participant experienced according to the 44-class solution. Adding participants' context count did not shift the estimates for these variance components, and this variable was not consistently significant. This indicates that diversity in contexts experienced did not systematically shift trait ratings.

Analysis 2

Which Contexts Are Important for Predicting Face Impressions?

Analytic approach. Next, we turned to a predictive modeling approach to assess which specific perceiver-level contextual variables might drive face impressions. We built a separate model for each of the 6 traits, where ratings on a trait (e.g., trustworthiness) served as the outcome variable in a cross-classified multilevel model, with each questionnaire item (e.g., "How sunny is it?") entered as a separate predictor. Models were cross-classified at the perceiver and target levels.

We had anticipated some multicollinearity among our numerous contextual variables and performed LASSO variable selection (Tibshirani, 1996) by incorporating L1-penalized estimation into generalized linear mixed-effects models to identify a more parsimonious model. However, results suggested we retain all variables in all models.

Accordingly, we built 6 models to predict ratings on each trait. We entered all 22 participant-mean centered contextual variables into the model (at Level 1) along with each participant's mean for each variable (at Level 2) to estimate both between- and within-perceiver effects. Models included random slopes for all level-1 predictors. Given the already complex model and no theoretically derived predictions, we did not include higher-order interactions (i.e., given 22 predictors, to estimate all three-way and two-way interactions would require estimating an additional 1,771 parameters). Models were estimated using the *lme4* (Bates et al., 2015) and *brms* (Bürkner, 2017) packages in R. See Supplementary Materials for code and further elaboration of this model.

Results. We investigated which specific perceiver-context predictors influenced impressions. Due to the large number of predictors and hypothesis tests, we interpreted effects as meaningful only if they were significant ($\alpha = .05$) across both exploratory and confirmatory datasets. See Supplementary Table 2 for comprehensive reporting of all within-perceiver and between-perceiver effects of contextual variables on each of the 6 traits.

Trustworthy and friendly. For impressions of both trustworthiness and friendliness ($r = .70$; typically highly correlated in impressions), only the between-perceiver effect of energetic mood was significant across both datasets. On average,

people who felt energetic more often than others judged faces as friendlier (exploratory: $\gamma_{010}\bar{X}_{energetic}$ = .38, 95% CI = [.16, .60], confirmatory: $\gamma_{010}\bar{X}_{energetic}$ = .26, 95% CI = [.05, .47]) and more trustworthy (exploratory: $\gamma_{010}\bar{X}_{energetic}$ = .32, 95% CI = [.10, .54], confirmatory: $\gamma_{010}\bar{X}_{energetic}$ = .22, 95% CI = [.01, .43]). No other effects were consistent across both exploratory and confirmatory sets.

Attractive and intelligent. Across exploratory and confirmatory datasets, none of the 22 contextual variables examined here had a consistent impact on ratings of attractiveness nor intelligence (r = .51).

Dominant and physically strong. For impressions of dominance and physical strength (r = .59; typically highly correlated in impressions), the between-perceiver effects of angry mood and hunger were significant across both datasets. On average, people who felt angry more often than others judged faces as less dominant (exploratory: $\gamma_{020}\bar{X}_{angry}$ = -.89, 95% CI = [-1.41, -.37], confirmatory: $\gamma_{020}\bar{X}_{angry}$ = -.47, 95% CI = [-.93, -.02]) and less physically strong (exploratory: $\gamma_{020}\bar{X}_{angry}$ = -.59, 95% CI = [-1.04, -.15], confirmatory: $\gamma_{020}\bar{X}_{angry}$ = -.43, 95% CI = [-.86, -.01]).

People who felt hungrier on average judged faces as more dominant (exploratory: $\gamma_{030}\bar{X}_{hungry}$ = .18, 95% CI = [.02, .34], confirmatory: $\gamma_{030}\bar{X}_{hungry}$ = .28, 95% CI = [.10, .46]) and physically stronger (exploratory: $\gamma_{030}\bar{X}_{hungry}$ = .16, 95% CI = [.03, .29], confirmatory: $\gamma_{030}\bar{X}_{hungry}$ = .22, 95% CI = [.05, .38]). Thus, participants who often felt angry judged faces as less dominant and physically weaker, whereas those who often felt hungry judged faces as more dominant and physically stronger.

Additional analyses. Though our theoretical interest centered on within-subject effects, to better characterize these between-participant effects, we explored whether participant gender or Big-Five personality scores were responsible for energetic mood, anger, and hunger effects. Specifically, we wanted to make sure these effects were robust even with other participant characteristics in the model.

The effect of energetic mood on ratings of trustworthiness and friendliness was *not* moderated by gender. Energetic mood remained significant even after controlling for Big-Five personality scores, indicating that participants who felt more energetic on average judged targets to be friendlier and more trustworthy—even after controlling for traits such as extraversion.

In addition, the effect of hunger on ratings of dominance and physical strength was *not* moderated by gender. Hunger remained significant even after controlling for Big-Five personality scores, indicating that participants who felt hungrier on average judged targets to be stronger and more dominant—even after controlling for traits such as agreeableness.

However, the effect of angry mood on ratings of dominance and physical strength—while not moderated by gender—was significant only in the exploratory dataset, when controlling for Big-Five personality scores. Descriptively, we found that participant-level angry mood was moderately correlated with Big-Five agreeableness (r = -.33) in the confirmatory dataset. This suggests that our measure of participants' average level of angry mood is related to participants' trait-level agreeableness, and may not explain enough variance in dominance and physical strength on its own.

Finally, none of the Big-Five personality dimensions examined here had a consistent significant effect on ratings along any of these six traits, across both exploratory and confirmatory datasets.

Sensitivity analysis. Throughout, we found no consistent effects of within-subject variation. One concern was that our power was too low to detect such effects, should they exist. Accordingly, we conducted a sensitivity analysis to determine at what power we would be able to detect any effects, given the variation observed in each variable. The power curve is available in Figure 5. Starting with a small effect at β = $\pm$ .20, we found that we had >99% power to detect this effect for 17 of our 22 variables, and >85% power to detect this effect for all 19 continuous variables of our 22 variables. For three variables (Are you with strangers? Are you with known others? Are you inside or outside?), we had much lower power. All three of the variables were dichotomous, and closer inspection revealed this result was likely due to low variance. For these three variables, our results should be viewed with caution. However, for the remainder, our within-subjects longitudinal design enabled high statistical power to detect within-subject contextual effects. Only for variables with a quite small true effect size of smaller than β = $\pm$ .10 would our tests be underpowered.

General Discussion

When scientists study impression formation in the lab, they typically want their findings to generalize to other contexts in which people form facial impressions. Yet research with greater external validity is difficult, limiting our ability to answer basic questions about social perception. Researchers studying impression formation have long considered that perceivers' contexts may be an important source of variance in impressions. Here, we present the first direct investigation into how people's day-to-day experiences shape their impressions. Using experience-sampling, we examined to what extent perceivers' daily experiences influenced the way that they form impressions from faces. Importantly, we found that perceivers' experienced contexts did not meaningfully impact their trait impressions. The average trait impression formed in one perceiver context did not differ from the average impression formed in another perceiver context. This suggests that certain perceiver-related factors (e.g., mood, environment, physiological state, psychological situation) are unlikely to shift their trait impressions of faces.
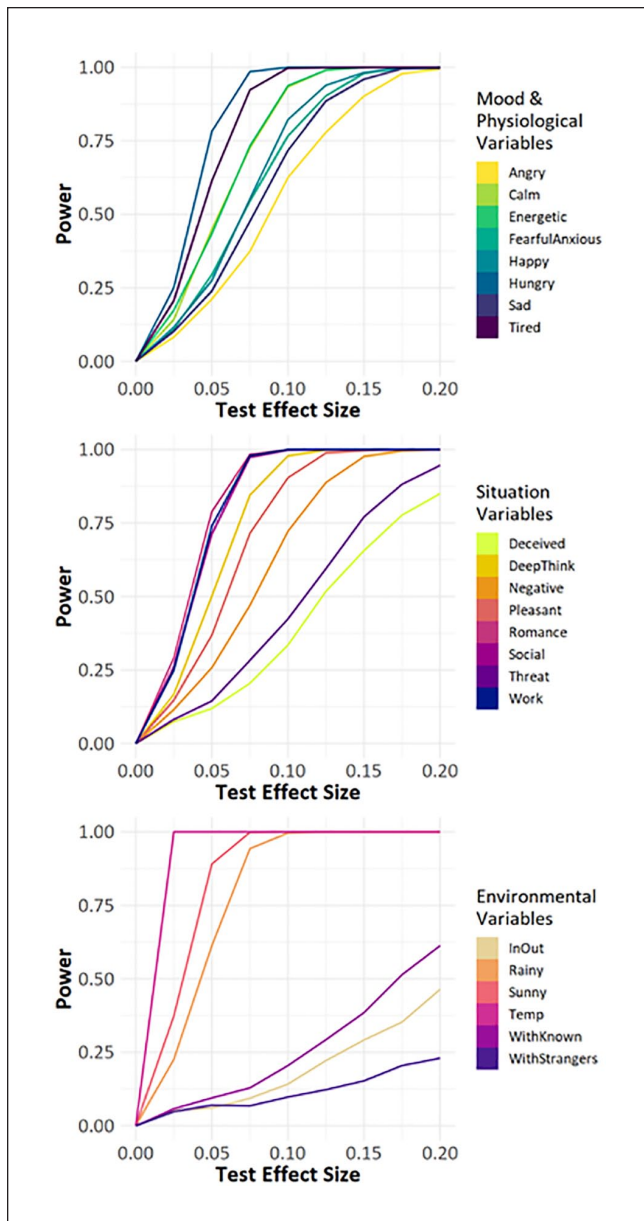
Figure 5. Sensitivity analysis demonstrating our power to detect various desired effect sizes, at $N_{participants} = 109$ (matched to the participant sample size of our exploratory/confirmatory datasets separately), $N_{stimuli} = 15$ (matched to the target sample size of our datasets), and $n = 2,236$ (to approximate our observed $n_{observations}$ at 2,236 and 2,190 across exploratory and confirmatory datasets).

Moreover, this conclusion did *not* vary across different perceivers experiencing different contexts. As a hypothetical example, if happy people judged others as friendlier on a sunny day whereas unhappy people did not, then the differences between perceivers (e.g., how happy they are) would interact with their experienced context (e.g., how sunny it is) to shape their impressions of targets. Yet, we found that the interaction between perceivers and their experienced contexts contributed a negligible amount (~1%) of the overall

variance in face impressions. To put this into perspective, recent work found that ~50% of between-perceiver variance can be attributed to positivity bias and an acquiescing response style (Heynicke et al., 2021; Rau et al., 2021). Our findings suggest that less than ~1% of this between-perceiver variance may additionally be accounted for by individuals' varying responsivity to their experienced contexts. Any differences in how individuals form trait impressions are unlikely to be driven by their experienced contexts, such as features of their local environment or psychological situation.

Our results converge with recent work that examines individual variance in how social impressions are formed from faces. For example, broader country-based cultural differences contribute negligible variance to face impressions relative to individual differences (Hester et al., 2021). Similarly, genetics seem to have little impact on impressions relative to one's personal upbringing and environment (Sutherland et al., 2020). Here, we investigated which of these individual differences matter more (i.e., by disentangling the contributions of stable individual differences from perceivers' situational, experienced contexts). We found that perceivers' day-to-day experienced contexts are unlikely to impact how they form impressions of others—and highlight the importance of other perceiver characteristics (e.g., personality or development) in shaping social perception.

Consistent with this interpretation, Analysis 2 found stable differences across perceivers in how they form impressions. Across multiple timepoints, participants who reported feeling more energetic than others judged targets as friendlier and more trustworthy. Those who reported feeling hungrier and less angry than others judged targets as more dominant and physically strong. These effects were independently produced four total times, across exploratory and confirmatory sets of two highly correlated traits (Xie et al., 2019; Zebrowitz et al., 2003). Since these mood and physiological perceptions were significant between-perceiver but not within-perceiver predictors, they represent individual differences and not within-person change over time. We discuss these between-person differences in a later section. Overall, the real-world contexts examined here do not meaningfully affect face impressions.

Consistent with previous work, we found that the perceiver-by-target interaction was by far the largest contributor to variance in facial impressions. Across multiple traits, estimates of perceiver-by-target contributions were similar but slightly smaller (20%–36%) than those in previous studies (32%–40%; Hehman et al., 2017; Hönekopp, 2006). The use of fewer stimuli compared to previous studies may have limited the variance in "personal taste" that could be captured by the perceiver-by-target interaction. However, 95% confidence intervals around ICCs for attractiveness, dominance, and physical strength contained the estimates obtained in previous work, providing evidence that these estimates generalize across multiple evaluative contexts (e.g., rating faces

in-lab or more naturalistically on a phone app) and study characteristics (e.g., rating one face vs. many faces per session, one trait vs. multiple traits at a time).

The variance uniquely attributable to perceiver characteristics alone was ~20% across traits, similar to previous work (20%–25%; Hehman et al., 2017; Xie et al., 2019). Thus, the inclusion of perceivers' experienced context did not partition out any meaningful variance in "idiosyncratic" or perceiver-level variability. This affects the interpretation of these clusters in cross-classified multilevel models, which are increasingly used in research on interpersonal judgments. Specifically, by partitioning the perceiver-by-context interaction, we can be more confident that what remains of "perceiver-level variance" in most lab-based studies of impression formation is specific to individual differences across people.

Finally, target characteristics uniquely contributed ~17% to the variance in facial impressions. This percentage can be interpreted as consensus (across perceivers) in trait impressions that are driven by differences in target stimuli. Given the focus of the present research on within-participant variability and the large number of questions, we purposely limited the number of target stimuli. Yet results from this smaller target set are consistent with estimates of target variance from previous research with much larger sets (10%–15%; Hehman et al., 2017; Xie et al., 2019), providing evidence that any results we obtained here were not a function of a smaller target set.

Overall, the residual unexplained variance was as large as 20% to 40% of the variance in previous research (Hehman et al., 2017; Xie et al., 2019). Our attempts here to incorporate perceiver context did not significantly reduce this unexplained percentage, as perceivers' experienced contexts do not seem to exert a strong influence on impressions. Critically, there are two practical implications of this work for future research on face perception. Researchers interested in examining sources of variance in trait impressions might be better served by investigating more stable individual differences, versus momentary situational factors experienced by the participant. Further, our results suggest that conclusions from face impression research conducted in lab or office settings may be likely to generalize to other perceivers' experienced contexts, though further research is required.

Participant Trait-Level Predictors

Though our theoretical focus was on within-person variation, we did find three between-person predictors of various trait impressions: anger, hunger, and energetic mood. While to our knowledge, these relationships have not been previously documented, they are consistent with some findings in related domains. For example, participants with higher average levels of anger rated targets lower on strength and dominance, consistent with functional accounts finding that anger was associated with lower perceptions of risk (Lerner & Keltner, 2000).

Similarly, participants with greater average hunger rated targets as stronger and more dominant. Previous work has found that people who are physically incapacitated perceive targets as larger and more muscular (Fessler & Holbrook, 2013). This conclusion is consistent with the present work to the extent that hunger correlates with feelings of weakness. Participants who feel hungrier on average may feel physically disadvantaged, and overestimate risk by perceiving targets as more dominant and formidable.

Finally, in novel social situations or when interacting with a stranger, energetic mood is characterized by a heightened tendency to approach positive stimuli (Elliot, 2006). Individuals who, on average, experienced higher energetic mood rated targets as friendlier and more trustworthy, with no impact on other traits—suggesting it was associated not with overall positivity, but with impressions relevant to approach appraisals.

Limitations

The present research was more externally valid than previous lab-based studies. Because participants were going about their day, any impressions formed of targets would better approximate the psychological contexts that scientists are hoping to capture in their research. Yet despite some advantages, the present design is still divorced from reality in some ways. Targets to be evaluated were still static and presented on a screen, and were not encountered naturally in the wild. Stimuli were contextually and emotionally neutral. We adopted this design intentionally to incrementally isolate one novel component of the day-to-day impression formation process (i.e., perceivers' experienced contexts), yet future research can continue to expand the external validity of impression formation research. Furthermore, while dynamic in-person evaluations are not captured here, people do regularly evaluate others from static photographs (e.g., dating apps, social media) in which targets are embedded in different contexts. For instance, target contexts such as visual scenery and the presence of other people can influence judgments of trustworthiness (Brambilla et al., 2018; Mattavelli et al., 2021), emotion (Barrett & Kensinger, 2010), and attractiveness (Carragher et al., 2021). While the present work explores perceiver contexts, more work is necessary to understand how perceiver and target contexts interact to shape impressions.

Second, the present work operationalizes perceivers' "day-to-day contexts" as a limited combination of environmental features, mood, physiological states, and psychological situations that were somewhat subjectively chosen by the researchers. To the extent that other perceiver contexts meaningfully impact impression formation, our estimates of contextual influence will be underestimates. Our conclusions are limited to perceiver contexts in which participants are able to complete a study on their phone. Responding to a survey on their phone may have momentarily removed perceivers from their

experienced context. Moreover, this design may limit the identification of specific contexts in which participants are unable or unwilling to respond to their phone. This may have contributed to low variance in the three categorical variables that had low power in our study (Are you with strangers? Are you with known others? Are you inside or outside?). Results for these three variables should therefore be viewed with caution. Although these contexts do not capture the range of all possible perceiver contexts, we have sampled regularly-experienced, day-to-day contexts. Future work could explore whether other (e.g., extreme, unusual) perceiver contexts reveal meaningful variation in impressions not captured here.

Finally, the longitudinal design necessitated a trade-off between comprehensiveness in our measures and minimizing participant fatigue to maximize response rate as they went about their day. The limited stimulus set used does not represent the diverse population of individuals who evaluated them, and future research might explore whether these contextual influences hold for different, more diverse, and less controlled stimuli. Although we used fewer stimuli than is typically reported in previous research, the present work focuses on perceiver context effects—and we do not expect our estimates of these effects to be biased by the limited number of stimuli. For example, across all 6 traits, our estimates of the perceiver-by-target ICCs, perceiver-ICCs, and target-ICCs replicate those reported in other studies with much larger (and more diverse) stimuli sets (e.g., ~800; Hehman et al., 2017; Xie et al., 2019). The correlation of trait ratings across timepoints ($r = .66$ in exploratory dataset, $r = .61$ in confirmatory dataset) was similar to those observed in datasets with more stimuli ($r = .72$; Hehman et al., 2017). We did not include any target-level predictors (e.g., target race, target gender) in our model, given low power to detect target-level effects. Finally, the use of a small, controlled set of context-neutral stimuli may have helped isolate any observed intraindividual variance to perceiver factors.

Conclusion

Impression formation researchers have long considered that perceivers' experienced context might be a meaningful source of variation in impressions. The present work contributes by testing this possibility, finding limited evidence that perceivers' contexts are an important factor in impressions. Perceiver context alone does not systematically influence trait impressions in a consistent manner—suggesting that perceiver and target idiosyncrasies are the most powerful drivers of social impressions. Importantly, we found no evidence to suggest that perceivers' experienced contexts could shape face impressions in a systematic way. This result tentatively suggests that the conclusions drawn from most social psychology research on impression formation, in which participants are seated in front of a computer, may be robust to fluctuations in day-to-day perceiver experiences of mood, environment, and perceived psychological situation.

ORCID iDs

Sally Y. Xie 🆔 https://orcid.org/0000-0002-1200-9470
Sabrina Thai 🆔 https://orcid.org/0000-0003-1333-6825

Supplemental Material

Supplemental material is available online with this article.

Note

1. R code:
   lmer, TraitRating ~ 1 + (1|PerceiverID) + (1|TargetID) + (1|ContextID) + (1|PerceiverID: TargetID) + (1|ContextID: PerceiverID) + (1|ContextID: TargetID), data)

References

Abele, A., & Petzold, P. (1994). How does mood operate in an impression formation task? An information integration approach. *European Journal of Social Psychology*, *24*(1), 173–187.

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46), 17948–17953. https://doi.org/10.1073/pnas.0705435104

Barrett, L. F., & Kensinger, E. A. (2010). Context is routinely encoded during emotion perception. *Psychological Science*, *21*(4), 595–599.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*(3), 338–363.

Bjornsdottir, R. T., & Rule, N. O. (2017). The visibility of social class from facial cues. *Journal of Personality and Social Psychology*, *113*(4), 530–546

Blair, I. V. Y., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science*, *15*(10), 674–679.

Brambilla, M., Biella, M., & Freeman, J. B. (2018). The influence of visual context on the evaluation of facial trustworthiness. *Journal of Experimental Social Psychology*, *78*, 34–42.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer, Jr. (Eds.), *Advances in social cognition* (*Vol. 1*, pp. 1–36). Lawrence Erlbaum. https://doi.org/doi:10.1521/soco.1993.11.1.150

Brown, N. A., Neel, R., & Sherman, R. A. (2015). Measuring the evolutionarily important goals of situations: Situational affordances for adaptive problems. *Evolutionary Psychology*, *13*(3), 1–15.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Carragher, D. J., Thomas, N. A., & Nicholls, M. E. (2021). The dissociable influence of social context on judgements of facial attractiveness and trustworthiness. *British Journal of Psychology*, *112*, 902–933.

Chartrand, T. L., van Baaren, R. B., & Bargh, J. A. (2006). Linking automatic evaluation to mood and information processing style: consequences for experienced affect, impression formation, and stereotyping. *Journal of Experimental Psychology: General, 135*(1), 70.

Duarte, J., Siegel, S., & Young, L. (2012). Trust and credit: The role of appearance in peer-to-peer lending. *The Review of Financial Studies*, *25*(8), 2455–2484.

Elliot, A. J. (2006). The hierarchical model of approach-avoidance motivation. *Motivation and Emotion*, *30*(2), 111–116.

Fessler, D. M. T., & Holbrook, C. (2013). Bound to lose: Physical incapacitation increases the conceptualized size of an antagonist in men. *PLOS ONE*, *8*(8), Article e71306. https://doi.org/10.1371/journal.pone.0071306

Fessler, D. M. T., Holbrook, C., & Snyder, J. K. (2012). Weapons make the man (larger): Formidability is represented as size and strength in humans. *PLOS ONE*, *7*(4), Article e32751.

Forgas, J. P. (1992). On mood and peculiar people: Affect and person typicality in impression formation. *Journal of Personality and Social Psychology*, *62*(5), 863–875.

Forgas, J. P., & Bower, G. H. (1987). Mood effects on person-perception judgments. *Journal of Personality and Social Psychology*, *53*(1), 53–60.

Freeman, J. B., Stolier, R. M., & Brooks, J. A. (2020). Dynamic interactive theory as a domain-general account of social perception. In B. Gawronski (Ed.), *Advances in experimental social psychology* (Vol. 61, pp. 237–287). Elsevier.

Hehman, E., Stolier, R. M., Freeman, J. B., Flake, J. K., & Xie, S. Y. (2018). Toward a comprehensive model of face impressions: What we know, what we do not, and paths forward. *Social and Personality Psychology Compass*, *13*(2), Article e12431. https://doi.org/10.1111/spc3.12431

Hehman, E., Sutherland, C. A. M., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology*, *113*(4), 513–529.

Hester, N., Xie, S. Y., & Hehman, E. (2021). Little between-region and between-country variance when forming impressions of others. *Psychological Science*, *32*, 1907–1917.

Heynicke, M., Rau, R., Leising, D., Wessels, N., & Wiedenroth, A. (2021). Perceiver effects in person perception reflect acquiescence, positivity, and trait-specific content: Evidence From a large-scale replication study. *Social Psychological and Personality Science*. Advance online publication. https://doi.org/10.1177/19485506211039101

Holbrook, C., Galperin, A., Fessler, D. M. T., Johnson, K. L., Bryant, G. A., & Haselton, M. G. (2014). If looks could kill: Anger attributions are intensified by affordances for doing harm. *Emotion*, *14*(3), 455–461.

Hönekopp, J. (2006). Once more: Is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *Journal of Experimental Psychology. Human Perception and Performance*, *32*(2), 199–209. https://doi.org/10.1037/0096-1523.32.2.199

Horstmann, K. T., Rauthmann, J. F., & Sherman, R. A. (2017). Measurement of situational influences. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *The SAGE handbook of personality and individual differences* (pp. 465–484). SAGE.

Hox, J. J., & Roberts, J. K. (2011). *Handbook of advanced multi-level analysis*. Psychology Press.

Jaeger, B., Evans, A., & van Beest, I. (2019). *Facial appearances and electoral success: Does regional corruption moderate preferences for trustworthy-looking politicians?* https://www.semanticscholar.org/paper/Facial-appearances-and-electoral-success%3A-Does-for-Jaeger-Evans/4a0e78a7b2a0ac37ff10d5bd947c46c10d7259a9

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. https://doi.org/10.1037/a0028347

Kawakami, K., Amodio, D. M., & Hugenberg, K. (2017). Intergroup perception and cognition: An integrative framework for understanding the causes and consequences of social categorization. In J. Olson (Ed.), *Advances in experimental social psychology* (1st ed., *Vol. 55*, pp. 1–80). Elsevier. https://doi.org/10.1016/bs.aesp.2016.10.001

Kenny, D. A. (2019). *Interpersonal perception: The foundation of social relationships*. Guilford Publications.

Kunda, Z., Thagard, P., Kunda, Z., Thagard, P., Kintsch, W., & Ross, M. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, *103*(2), 284–308.

Lerner, J. S., & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition & Emotion*, *14*(4), 473–493.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*, 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*, 93–120. https://doi.org/10.1146/annurev.psych.51.1.93

Maner, J. K., Kenrick, D. T., Becker, D. V., Robertson, T. E., Hofer, B., Neuberg, S. L., Delton, A. W., Butner, J., & Schaller, M. (2005). Functional projection: How fundamental social motives can bias interpersonal perception. *Journal of Personality and Social Psychology*, *88*(1), 63–78.

Mattavelli, S., Masi, M., & Brambilla, M. (2022). Not just about faces in context: Face–context relation moderates the impact of contextual threat on facial trustworthiness. *Personality and Social Psychology Bulletin*. Advance online publication. https://doi.org/10.1177/01461672211065933

Matthews, G., Jones, D. M., & Chamberlain, A. G. (1990). Refining the measurement of mood: The UWIST mood adjective checklist. *British Journal of Psychology*, *81*(1), 17–42.

Merz, E. L., & Roesch, S. C. (2011). A latent profile analysis of the Five Factor Model of personality: Modeling trait interactions. *Personality and Individual Differences*, *51*(8), 915–919.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569.

Oh, D., Dotsch, R., Porter, J., & Todorov, A. (2019). Gender biases in impressions from faces: Empirical studies and computational models. *Journal of Experimental Psychology: General*, *149*, 332–342.

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior*, *34*(2), 83–110. https://doi.org/10.1007/s10919-009-0082-1

Parrigon, S., Woo, S. E., Tay, L., & Wang, T. (2017). CAPTIONing the situation: A lexically-derived taxonomy of psychological situation characteristics. *Journal of Personality and Social Psychology*, *112*(4), 642–681.

Pastor, D. A., Barron, K. E., Miller, B. J., & Davis, S. L. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemporary Educational Psychology*, *32*(1), 8–47.

Peugh, J., & Fan, X. (2013). Modeling unobserved heterogeneity using latent profile analysis: A Monte Carlo simulation. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(4), 616–639.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 111–163.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212.

Rau, R., Carlson, E. N., Back, M. D., Barranti, M., Gebauer, J. E., Human, L. J., Leising, D., & Nestler, S. (2021). What is the structure of perceiver effects? On the importance of global positivity and trait-specificity across personality domains and judgment contexts. *Journal of Personality and Social Psychology*, *3*(120), 745–764.

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nav, C. S., Sherman, R. A., Ziegler, M., Jones, A. B., & Funder, D. C. (2014). The situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, *107*(4), 677–718. https://doi.org/10.1037/a0037250

Rauthmann, J. F., & Sherman, R. A. (2015). Ultra-brief measures for the situational eight DIAMONDS domains. *European Journal of Psychological Assessment*, *32*, 165–174.

Rauthmann, J. F., & Sherman, R. A. (2018). The description of situations: Towards replicable domains of psychological situation characteristics. *Journal of Personality and Social Psychology*, *114*(3), 482–488. https://doi.org/10.1037/pspp0000162

Rosenberg, J., Beymer, P., Anderson, D., van Lissa, C. J., & Schmidt, J. (2019). TidyLPA: An R package to easily carry out Latent Profile Analysis (LPA) using open-source or commercial software. *Journal of Open Source Software*, *4*(44), 978.

Stolier, R. M., Hehman, E., & Freeman, J. B. (2020). Trait knowledge forms a common structure across social cognition. *Nature Human Behaviour*, *4*(4), 361–371.

Stolier, R. M., Hehman, E., Keller, M. D., Walker, M., & Freeman, J. B. (2018). The conceptual structure of face impressions. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(37), 9210–9215. https://doi.org/10.1073/pnas.1807222115

Sutherland, C. A. M., Burton, N. S., Wilmer, J. B., Blokland, G. A. M., Germine, L., Palermo, R., Collova, J. R., & Rhodes, G. (2020). Individual differences in trust evaluations are shaped mostly by environments, not genes. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(19), 10218–10224.

Thai, S., & Page-Gould, E. (2018). ExperienceSampler: An open-source scaffold for building Smartphone apps for experience sampling. *Psychological Methods*, *23*(4), 729–739.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance. *Annual Review of Psychology*, *66*(15), 1–27. https://doi.org/10.1146/annurev-psych-113011-143831

Wilson, J. P., Hugenberg, K., & Rule, N. O. (2017). Racial bias in judgments of physical size and formidability: From size to threat racial bias in judgments of physical size and formidability: From size to threat. *Journal of Personality and Social Psychology*, *113*(1), 59–80. https://doi.org/10.1037/pspi0000092

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science*, *26*(8), 1325–1331. https://doi.org/10.1177/0956797615590992

Xie, S. Y., Flake, J. K., & Hehman, E. (2019). Perceiver and target characteristics contribute to impression formation differently across race and gender. *Journal of Personality and Social Psychology*, *117*(2), 364–385.

Xie, S. Y., Flake, J. K., Stolier, R. M., Freeman, J. B., & Hehman, E. (2021). Facial impressions are predicted by the structure of group stereotypes. *Psychological Science*, *32*(12), 1979–1993. https://doi.org/10.1177/09567976211024259

Zebrowitz, L. A., Fellous, J.-M., Mignault, A., & Andreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: Evidence from connectionist modeling. *Personality and Social Psychology Review*, *7*(3), 194–215.