






Social Psychological Bulletin

Beyond Reliability in First Impressions Research: Considering Validity and the Need to “Mix It Up With Folks”

Liam Satchell¹ , Bastian Jaeger² , Alex Jones³ , Beatriz López⁴ ,
Christoph Schild⁵ 

[1] *Department of Psychology, University of Winchester, Winchester, United Kingdom.* [2] *Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, Amsterdam, Netherlands.* [3] *School of Psychology, Swansea University, Swansea, United Kingdom.* [4] *Department of Psychology, University of Portsmouth, Portsmouth, United Kingdom.* [5] *Department of Psychology, University of Siegen, Siegen, Germany.*

Social Psychological Bulletin, 2023, Vol. 18, Article e10211, <https://doi.org/10.32872/spb.10211>

Received: 2022-09-05 • **Accepted:** 2023-02-18 • **Published (VoR):** 2023-11-17



Handling Editors: Simine Vazire, Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, Australia; Brian Nosek, University of Virginia, Charlottesville, VA, USA

Corresponding Author: Liam Satchell, University of Winchester, Sparkford Road, Winchester, Hampshire, S022 4NR, United Kingdom. E-mail: liam.satchell@winchester.ac.uk

Related: This article is part of the SPB Special Topic "Is Psychology Self-Correcting? Reflections on the Credibility Revolution in Social and Personality Psychology", Guest Editors: Simine Vazire & Brian Nosek, Social Psychological Bulletin, 18, <https://doi.org/10.32872/spb.v18>

Abstract

‘First impressions’ are a popular topic in social psychology. They are researched because the initial judgments of others are consequential in everyday life (such as job interviews, first dates, justice outcomes). In the context of broader concerns about the credibility of psychological science, first impressions research has developed commendable initiatives for improving reliability (open stimulus databases, international collaborations, replication studies and reanalyses). However, these initiatives can impact the validity of studying how people form first impressions. There is a long history of critiquing the usefulness of passive-observer judgments of controlled, reduced, presentations of people—and these concerns are still relevant today. Here, we highlight the praiseworthy practices improving reliability in first impressions research, before moving on to identify persistent methodological concerns in the field. This includes inadequate stimulus sampling and diversity, constrained participant response options, limited consideration of study context, and limitations of atomised presentations of target people. We identify how these



This is an open access article distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), CC BY 4.0, which permits unrestricted use, distribution, and reproduction, provided the original work is properly cited.

methodological limitations impact theory development, how we might be over/underestimating everyday experience, and even misunderstanding social differences in autism and mental health. Finally, we identify opportunities for methodological reform, focusing on codifying instead of controlling interactions, promoting inductive, participant-led, methodologies, and asking for stronger theory development and clarity on ‘can’ vs. ‘do’ research questions. Overall, we praise reforms for improving the reliability of first impressions research, but improvements to making scientific predictions about first impressions require renewed consideration of validity.

Keywords

first impressions, person perception, replication crisis, validity, reductionism

Highlights

- First impressions research has responded to the replication crisis in many ways, including making stimulus sets open, sharing data for re-analysis, replication studies, and worldwide collaborations.
- However, the validity of common paradigms might be questioned given the asocial nature of participants observing and forced-responding to restricted, atomised, stimuli.
- We need to consider validity in first impressions as we lack models which can be readily applied to everyday experiences—and we may even be misunderstanding social differences such as in autism and mental health.
- We encourage researchers to consider validity and ask if they are studying ‘can’ or ‘do’ questions.

The topic of first impressions¹ is a popular area of study in social psychology. Understanding what influences initial perceptions of people is of everyday practical (job interviews, security evaluation), social justice (prejudicial biases), and personal (dating, romantic relationships) importance. Most studies on this topic examine first impressions based on images of faces or bodies, or recordings of voices presented for very brief periods, sometimes for less than 100 milliseconds (Willis & Todorov, 2006). However, many interpersonal experiences last longer than a second, and there are questions about the validity of reduced, briefly presented, static presentations of other people for understanding the complex contexts described above. The information richness that comes from dynamic interpersonal experiences means that first impressions can change

1) The term ‘first impressions’ can be interpreted differently. Here we mean the resultant judgments or perception (*impression*) of a person from a *first* exposure to another person. This may also be referred to as impression formation, person judgment, face/voice perception, or social cognition research by others in the field. This includes attempts to study pre-conscious judgments of a person (first-most impressions) which lend themselves to different methodologies not addressed here.

—for better or worse—depending on how the interaction plays out. This process is less well understood. Research comparing perceptions of others from photographs with the post-interaction judgments finds that first impressions can shift to be more positive after just five minutes of interaction (Satchell, 2019). Methodologies that use naturalistic first interactions are rare in the person perception literature. These studies allow participants to interact for the first time, with no prior knowledge of each other, and come to first impression judgments. For example, round-robin designs, where participants' first impressions are recorded after getting acquainted for a few minutes or even hours, have been used to study first impression accuracy, meta-accuracy (i.e., whether people know how others see them), and related phenomena (e.g. Albright et al., 1988; Tissera et al., 2021). One possible reason for why interaction studies are less common is because they involve less control over the stimulus (i.e., the target people). Obtaining large samples of diverse interacting partners can be difficult, and the exact study design can be seen as less reproducible. Doing interactive research in a way that emulates everyday contexts leads to reduced experimental control over participants' makeup, hair, clothing, mood, emotional expressions, conversation topics, and many more aspects of in vivo interaction.

There are clear benefits to using standardised photographs or voice recordings of target people given the wider reliability (reproducibility, replicability) crisis in social psychology. Here, we identify how methodological changes to improve the reproducibility and replicability of first impressions research may come at the cost of their *validity*. As others have noted in discussing the 'mutual-internal-validity' problem, fields like first impressions can become overly focused on iteratively understanding methodological elements (different ways to present a face or an eye), instead of taking diverse approaches to understand the target phenomenon—how people might experience their first interactions (H. Lin et al., 2021). We suggest that more research needs to focus on the validity of first impressions methodologies and how we can develop and analyse naturalistic contexts in a reproducible manner.

We are not the first to raise concerns about the asocial nature of person perception research. In 1980, observing the research of his contemporaries on 'social knowing', Ulric Neisser noted that the field overwhelmingly examined social perceptions by using passive observers. Participants were presented with atomised (broken down to core elements) stimuli such as photographs of faces or simple outlines of people. Neisser noted that the participant in these studies "doesn't mix it up with the folks [they're] watching, never tests [their] judgments in action or interaction" (Neisser, 1980, p. 603). In the last 40 years, arguably little has changed in the methodology used for studying first impressions. Participants are still routinely passive observers of atomised presentations of people with forced choice response options (by scale or categorisation). Whilst there have been improvements in terms of participant and stimulus sampling, and analytic considerations (see discussion below), the core studied task looks the same. For example,

88–90% of studies on ‘faces’ published in top social psychology journals use static images of faces in passive observer paradigms (Dawel et al., 2022). There are questions to be asked regarding how much our knowledge of first impressions has advanced since Neisser’s expression of concern. Especially when considering a reasonable scientific goal of trying to make a prediction about how a particular person would be perceived in a first, real, interaction with an unknown person.

Our paper first highlights recent commendable developments aimed at improving the reproducibility and replicability of first impressions research. Then, we highlight persistent methodological concerns with common practices and paradigms which might limit our understanding. We go on to describe how methodologies to enhance reliability may come at the cost of theory building and we highlight existing and promising alternative approaches which enable methodological robustness, whilst allowing us to get closer to our target phenomenon. This includes the use of techniques like lens modelling to codify rather than control interpersonal research and participant-led approaches. These are currently used in some research in this area, and we argue they should be adopted more broadly.

Good Reliability Practices in First Impressions Research

There are a range of ways in which first impressions research has developed good practice in response to the wider reproducibility crisis. Whilst there has been less published evidence of a replication crisis specifically in first impressions research, the principles of openness, reproducibility, and replication have been promoted in many recent studies in this field. These include open-access stimulus databases, reanalyses and replications, and open international collaborations.

Open Stimulus Databases for Reproducibility

One way to test the replicability and improve the accessibility of first impressions research is through sharing open stimuli. This can address the challenges that individual laboratory groups might have with the demands of producing large databases of controlled stimuli. Open-access databases can enable and democratise research by providing opportunities for smaller research groups with fewer resources to conduct high-quality studies.

Examples of good practice in this area include large open databases of faces (i.e. the ‘Chicago Face Database’; Ma et al., 2015) and voices (i.e. the ‘Jena Speaker Set’; Zäske et al., 2020). Such databases are of broad use; the Radboud Face Database (Langner et al., 2010), published in 2010, has been cited 2329 times at the time of writing. Making such stimulus databases open allows for collaborative expansions on these stimuli, such

as with the later inclusion of multiracial targets (Ma et al., 2015). Additionally, other researchers can collaboratively provide more data on the stimuli themselves for the benefit of all those who use the set. For example, facial landmark templates for all images were later added to the Chicago Face Database (Singh et al., 2022) and ratings of targets on 19 trait dimensions are available for the Radboud Face Database (Jaeger, 2020).

Large International Collaborations

One of the most notable consequences of the credibility crisis in psychology is the development of large collaborative initiatives like the Psychological Science Accelerator (PSA). The PSA is a ‘distributed network of laboratories’ dedicated to coordinating large-scale research projects (Moshontz et al., 2018). The development of the PSA was motivated by using crowdsourcing techniques “to accelerate the accumulation of reliable and generalizable evidence in psychological science” (Moshontz et al., 2018, p. 503). Inspired by other ‘Many Labs’ projects which coordinate large-scale, crowdsourced, replications, the PSA aims to collect data from a diverse range of countries and contexts. The first project tested by the PSA collaboration was a first impressions project. They investigated the generalizability of the valence-dominance model of first impressions of faces (developed by Oosterhof & Todorov, 2008) across cultures (Jones et al., 2021). The multinational collaboration used an ethnically diverse set of faces and sampled 11,570 participants from 41 countries in 11 world regions. This work was an impressive undertaking and a strong response to the concerns that psychological research, as a whole, faces about its credibility. This project is a model of how to consider the complexities of international work with substantial effort dedicated to translation, ethics, and stimulus diversity. Whilst other PSA projects focus on other areas of psychology, it is a good sign for the field of first impressions that a project on rating photographs of faces was the test case project.

Replication and Reanalysis

The availability of study materials, open data, and the willingness to facilitate others’ replications of published findings is an important part of developing a more credible science. In recent years, many first impression researchers have attempted to replicate and extend published findings to test the robustness of claims (e.g., Caton et al., 2022; Kramer & Gardner, 2020). For example, Ert and colleagues (2016) found that more trustworthy-looking Airbnb hosts in Stockholm are able to charge higher prices for qualitatively similar apartments, presumably because they are favoured by guests. Jaeger and colleagues (2019) conducted a conceptual replication by analysing a larger sample of Airbnb hosts from New York City while also controlling for additional characteristics of the apartment or the host that could confound the relationship between facial appearance and prices.

Replication efforts have also been facilitated by an increasing focus on data sharing. The first PSA project (Jones et al., 2021) produced a large and rich data set on first impressions of faces that has already been used to test unrelated hypotheses in follow-up investigations. For example, Batres and Shiramizu (2022) used the data to show that the attractiveness ‘halo effect’ (association of more positive traits with more attractive individuals) not only emerges reliably in Western samples, which were primarily used in previous studies, but also in other cultures. These examples highlight that shared data and stimuli can be enormously helpful for a field.

Persistent Methodological Concerns

While there are gains in replicability with the use of standardised shared stimuli, there are also potential costs. The focus on standardization and experimental control means that most studies assess constrained responses (e.g., in two-alternative forced-choice designs) to simplified stimuli, lacking in diversity, which are stripped of most, if not all, contextual factors that characterise everyday social interactions.

Inadequate Stimulus Sampling and Diversity

Conversations about adequate stimulus sampling have a long history in social psychology (Brunswik, 1955). As Hammond noted in 1948: “psychologists maintain a one-sided emphasis on the need for representativeness. They emphasize representativeness of populations, but not situations, tests, or objects - thereby implicitly limiting the generalization of results obtained to the population, or subject, side.” (p. 531). Whilst there is no simple, constant, or adequate number of target people needed for person perception research (much like samples of participants, this can be addressed in a simulation-based power analysis—for a tutorial see Kumle et al., 2021) it is important for researchers to acknowledge that each individual target person brings important variability to perception data.

To this end, researchers promote the use of linear mixed models (and their variations) for social psychology researchers to better understand how much of an influence the random factors of stimulus person variability has on the target effects being studied. A key consideration in effectively understanding these variance components is the need to use larger numbers and more diverse samples of stimulus people for first impressions studies. It should also be noted that within-person variability (how a person appears across photographs, contexts, times) has an important impact on how they are perceived as well, and naturally varying presentations of each target person (a driver’s licence photograph versus a staff profile picture versus a social media profile picture) should be considered in stimulus databases.

There have been clear efforts to improve the ethnic, age, and gender diversity of stimulus people in studies (see the stimulus databases mentioned above) to try and expand the generalisability of findings. [Foo and colleagues \(2022\)](#) meta-analysed research on face-based trustworthiness impressions and noted that the majority of studies did not describe the ethnicity of the target faces or raters in their method sections. They note that there were “potential biases, including a preponderance of Western studies, a lack of “cross-talk” between research groups, and clarity issues.” More specifically [Cook and Over \(2021\)](#) highlighted how most first impressions research is focused on White faces. They identified two parallel literatures on perceptions of faces: a considerable literature on investigating elements of face shape variance within White faces, and separately research with non-White faces—which are typically studied in contrast to White faces. Cook and Over observed that first impression papers “typically offer little or no justification for their use of ethnically homogeneous White face stimuli” (p. 5). It is important that first impressions researchers justify their stimulus choice and consider general trends in social psychology to focus more on diversity and inclusion. We should also consider different dimension of diversity that are important, such as neurodiversity, sexuality diversity, and social diversity.

It should also be noted that the way in which researchers instruct their target people can also restrict their appearance to perceivers. For example, when considering facial expressions of emotion, [Barrett and colleagues \(2019\)](#) reviewed the emotion recognition literature and found that most studies use prototypical facial expressions of core emotions. These stimuli involve extreme forms of target facial expression from a range of emotion categories chosen by the researchers. People posing as stimuli in these studies are encouraged to model these atypical presentations of emotion. As [Durán and Fernández-Dols \(2021\)](#) noted in their meta-analysis, in expression-generation studies, the prototypical expressions do not reliably occur with the intended emotional experience. Such strong forms of emotion expression are not typical in everyday experience. Further, in an analysis of spontaneous emotion footage (from 6.1 million open source videos from YouTube), variance in emotional expression was better explained by 16 dimensions rather than six (e.g., [Cowen et al., 2021](#)). [Barrett and colleagues \(2019\)](#) provide a thorough review of the methodological limitations in judgments of facial expressions of emotion, but in general, this area of research provides a salient example of how researchers may constrain their stimuli due to theoretical models. A field norm of six primary emotions leads to coached, exaggerated presentations of people for perceiving, which may poorly represent naturally occurring diversity in emotional expression. A useful example of how stimulus creation can have an impact on our investigation of everyday contexts.

Constrained Participant Responses Options (2AFC and Rating Scales)

First impression studies not only constrain the types of stimuli participants are exposed to, but also how participants can respond to the stimuli. For example, first impressions are often measured using a two-alternative forced-choice (2AFC) design. This involves showing two versions of the same stimulus, such as faces, bodies, or voice recordings, next to each other, and perceivers are instructed to directly compare which version scores higher on some dimension of interest (e.g., [Alper et al., 2021](#)). This design, which we refer to as the ‘evil twin paradigm’, is often used to study a specific characteristic in impression formation. For example, two versions of face images are created where the facial width-to-height ratio (fWHR) is reduced in one version and increased in the other version ([Stirrat & Perrett, 2010](#)). Participants view the two face versions and choose which one appears more trustworthy. If participants identify the face versions with lower fWHR as more trustworthy, then this is taken as evidence that people rely on fWHR when forming trustworthiness impressions. However, this inference is problematic because it is possible that perceivers are forced to rely on the manipulated facial characteristic when this is the only cue they have at their disposal ([A. Jones & Jaeger, 2019](#)). This is not necessarily what they would rely on in everyday life when they also have access to an array of other cues. The paradigm also highlights differences in facial appearance that may go completely unnoticed in everyday impression formation. For instance, fWHR may only be related to trustworthiness impressions when targets’ gender, age, ethnicity, facial expression, and other features are kept constant (as is the case in evil twin studies), but not when these dimensions vary (as is the case in everyday life; [Jaeger & Jones, 2022](#)). In this way, constrained response options can introduce false positive results (see [Bovet et al., 2022](#)).

Likert-style rating scales allow participants to give more nuanced responses to stimuli, but they still constrain how participants can react to stimuli. That is, ratings that are assessed (e.g., trustworthiness impressions) may not reflect people’s dominant reaction to a person in everyday life when their response options are not constrained (e.g., “this seems like a friendly person” such as in [Sutherland et al., 2018](#)).

Limited Consideration of Context

The most common experimental design in first impressions research asks a participant to pay close attention to a series of faces, bodies, or voices, and to form a judgement of them. But this focused observation is very different to the noisiness of everyday life when observations of others might be incidental, in passing, or occur in specific contexts. It could well be the case that particular parts of a whole person (such as their eyes or height), which we might consider to be important from a theoretical standpoint, might turn out to be less important when studied in a complex context. Recent work

demonstrated that during the observation of a social interaction in naturalistic scenes, just 14% of participant eye gaze patterns were to the face of the social interactors (Varela et al., 2023). These results highlight how researchers' assumptions about aspects of people that we might consider to be highly relevant—such as faces, which receive significant dedicated study in first impressions—might not be highly relevant in a noisy everyday environment.

Researchers rarely sample situations to the same extent that they sample participants or stimuli. Representative design approaches have long noted that different environments should also be considered for the impact they have on study findings (Brunswick, 1955). More than treating the laboratory environment as a fixed neutral experience, it is an environment of specific constraints. One controlled environment might tell us less about first impressions compared to sampling across multiple natural contexts. These settings elicit unique, specific behaviours. For example, there might be more to learn from seeing which faces consistently draw the attention of a participant in a restaurant, in a lecture hall, and on a street compared to how they might form a first impression of a face on a computer in a quiet, neutral room. Sampling across contexts and quantifying their variability is of significant theoretical value when the alternative is passive observation in a fixed environment.

Relatedly, there is limited investigation of classic first impressions questions in contexts with sufficient stakes to understand how important these findings are in practice. Consider a job interview, where the applicants might present considerable variability in job experience, expertise, and qualifications—as well as appearance, gender, and ethnicity. Despite research suggesting that perceptions of competency and related factors are influenced by facial appearance, limited research has explored how these factors explain concrete decisions, such as hiring, in complex settings (c.f. papers such as Carlsson & Eriksson, 2019). Understanding the variance explained by first impressions aspects in job applications or interviews is an important step in making our research more relevant to target contexts.

Reductionism and Atomisation of Target People

There are two main reasons to critique reductionist or atomised stimuli in first impressions research. The first is from a statistical covariance perspective, where it is important to consider that parts of a person are likely correlated. The second is from a theory perspective where we recognise classic Gestalt perception theory and how wholes are not equal to the sum of their parts.

Correlated Cues

Many unjustified inferences about the importance of specific elements of a person for first impressions derive from the fact that many cues that could form the basis of people's impressions are correlated (Jaeger & Jones, 2022). This makes it difficult to

determine which cues perceivers (primarily) rely on when forming impressions. Dozens, if not hundreds, of studies have examined which facial characteristics people rely on in first impressions by testing for associations between specific characteristics (e.g., fWHR) and perceivers' first impressions. Significant associations between facial characteristics and impressions are often interpreted as evidence that perceivers attend to and rely on these characteristics in impression formation. However, in virtually all studies, only one or a few facial characteristics are tested at a time. It thus remains unclear if these associations arise because perceivers rely on the tested facial characteristic when forming impressions, or because perceivers rely on another facial characteristic that is correlated with the one that is tested. These confounds are often not addressed in the literature.

Wholes Are not Sums of Parts

It is well known that the perception of a whole is not simply the perceived sum of its parts, as was first argued nearly 100 years ago in the Gestalt psychology movement. However, the approach to understanding first impressions through a series of atomised presentations of a person implicitly argues that we can make sense of the context we are emulating in a lab (meeting a whole person) through studying their parts in isolation. That is, using the sum of literature on atomised parts to infer the perception of the whole. As such, the field is an array of unique “bubbles” of finding, and it is not clear if these parts of the literature will burst when they come into contact with each other (Satchell, 2019). We may have data on the relevance of voice pitch on perceptions of attractiveness, the effect of face shape on dominance ratings, and the relationship between gait biomechanics on threat judgments, but the extent to which all of these interact in the whole is a question first impressions researchers cannot answer at this time. How is a person with a medium-pitched voice, narrow face, and heavy shoulder swagger perceived?

The impact of different aspects of a person has rarely been tested simultaneously. A recent study (Jaeger & Jones, 2022) showed that when modelling first impressions of faces from a large array of theoretically-motivated predictors (such as emotion resemblance, ethnicity, and morphology) using regularised models, only a small subset of predictors were retained, whereas many other predictors, including popular measures, such as fWHR and eye size, were omitted from the model due to their low unique informativeness when predicting first impressions. This suggests that in a wider, noisier context of a whole face, these elements were less relevant. The results may differ even more when considering perceptions of a whole individual's appearance, movements, and reactions.

From what we know about perception, we should be sceptical that wholes are merely additive sums of parts. Bringing together large numbers of well-studied atomised parts of a person is no guarantee of explaining first impressions. Engaging more with the question of how interested we are in perceptions of *people* in the field of first impressions

is an important future trajectory to enhance the usefulness and quality of our work further.

Theory Development Consequences

The above methodological concerns highlight how we might be fundamentally limiting our understanding of our target phenomenon. Restricting how our participants can indicate their impressions of new people, limiting our stimuli appearance, number, and diversity, and avoiding context and intentions all lead to a poorer estimate of the nature of first impressions. Further, the lines of research which try to define individuals as having ‘disordered’ or ‘deficient’ social perception abilities (for example, in Autism or mental health conditions) use these same asocial methodologies. We may be misrepresenting the quality of ‘social skills’ through testing participants under specific constraints in which there is limited space for social skills (action and reaction with others). Restricted first impressions methodologies limit our theoretical advances for understanding the everyday experiences of typical populations as well as those studied with social differences.

Over/Underestimation of Everyday Experiences

The issue with methodologies that do not effectively emulate everyday experiences is that they might both over- and under-estimate the importance of key social features in everyday interaction. We would encourage readers to think beyond experimental control and replication, to the validity of the task in hand. For example, there may be parts of a person that are difficult to perceive or rarely attended to in noisy everyday situations, but these rise to prominence in experimental designs where they do affect first impressions when they are the only thing a participant sees. When participants only see a face or an eye, the information available to them to make a judgment is not a valid representation of the process of forming a first impression with a whole person before them. Eye colour might be important in a limited presentation of a person (where only the eye region is visible to the participant), but a person’s eye colour may have little influence on how we might form a first impression when that person is wearing an absurd hat or doing an unusual dance.

The relevance for atomised presentations of people in understanding noisy, information-rich, context-laden everyday situations, like a job interview, first date, or legal process, is extremely limited. However, the concern that laboratory research does not represent the effects in the wild is not just to claim that the effects might be artificially increased in experiments. It could well be the case that, within the range of all things going on in a high-stakes interaction, eye contact or face qualities or voice tone might play an *oversized* role in shaping first impressions. Perhaps an atypical voice outweighs the presence of an unusual hat in everyday life, but we can only know this through

studying this in the presence of competing sources of variance. In order to build stronger, functional, theory, we should consider the value of holistic experiences of other people.

Misunderstanding Social Differences in Autism and Mental Health

Excluding interpersonal dynamics from first impressions research can also alter our research on diagnoses characterised by social behaviour differences. When the materials and methods used to build models of social differences are not based on social interactions, we might misunderstand differential responses as ‘social deficits’. Many of these lines of research are motivated by noticeable everyday challenges that some people experience, yet the selection of methodologies used to understand these differences might exclude or be misleading about the nature of these difficulties.

For example, in the case of autism, evidence for the most influential theory—that autism is resultant from the theory of mind deficits—largely rests on a widely used test that requires participants to match researcher-chosen words to photographs of magazine models’ eyes presenting an emotion, the ‘mind in the eyes test’ (Baron-Cohen et al., 1997). Similarly, the more recent ‘reduced social motivation’ theory of autism (Chevallier et al., 2012) also rests on evidence showing reduced social attention in autism to faces and eye regions of photographs (e.g., Chita-Tegmark, 2016). Yet, there are questions about how observations of photographs are comparable to being social with others. We should note that the ‘mind in the eyes’ test suffers from replication and measurement issues (see Gernsbacher & Yergeau, 2019), but more importantly, there are questions about the extent to which atomised images of actors presenting an emotion represent an everyday ‘mind reading’ activity. Therefore, we should question how ‘low scores’ on this task correspond to a social deficit. Critiques of these methods are supported by more recent studies including interacting social partners, which consistently fail to find differences between autistic and non-autistic participants in social attention patterns (see Kikuchi et al., 2022). In fact, many of these studies show that the differences present in autism are the result of subtle interpersonal dynamics such as the topic of the conversation (e.g., Hutchins & Brien, 2016), rather than a general lack of social motivation.

This leads to broader questions about methods that use photographs or digital renders of faces to gain insight into social differences across human neurodiversity. A variety of meta-analyses and systematic reviews highlight how a common approach to understanding social challenges starts with presenting images of emotional states—for example, in research on antisocial/dissocial personality disorder and psychopathy (Marsden et al., 2019), social phobia and Post Traumatic Stress Disorder (Plana et al., 2014), and in attention, eating, and anxiety disorders in child psychiatry (Collin et al., 2013). Many of these reviews do find differences between diagnoses in responses to faces, and these differences are of note in trying to understand these topics. However, more work is needed to understand how face processing relates to the complexity of everyday social encounters—especially when many of these diagnoses are related to a variety of

dynamic interpersonal behaviours, not just facial image processing. It is important for future research to consider how much variance these aspects of first impressions matter for our studies of neurodiversity.

Opportunities and Future Directions

We have identified good practice in terms of improving reliability of first impressions research, perhaps at the cost of validity and phenomenon clarity. We now look at opportunities to continue to be open, robust, and reliable without compromising validity. This includes considering ways to enhance the diversity and quality of controlled paradigms, making naturalistic designs more acceptable in first impression research, giving more control to participants (and handling their contributions reproducibly), and further improvements to theory development.

Codifying, Not Controlling, Interactions

As mentioned above, the atomised stimulus person, presented without context or dynamics, may not represent what many might experience as a ‘first impression’. It is important to consider how these first encounters with others—especially in settings of high stakes like job interviews, first dates, and legal interactions—are shaped by the complex environment and context of the interaction. There are frameworks that enable science in this way. Egon Brunswik famously proposed ‘lens model’ approaches to perception activities (Brunswik, 1955) and this has been applied by some researchers of first impressions (i.e., Nestler & Back, 2013). In these frameworks, participants are invited to judge another person in a context. Then aspects of the ecology and perceptual targets are coded. The naturally occurring variability between targets in the ecology reflects the different ways the targets provide information for the perceiver. In a general example, one might consider trying to judge whether someone in a store is an employee or a fellow customer. The other people in the store will vary on many aspects, for example, wearing a uniform, who they are with, or their age. Judgments of whether someone is an employee will be a consequence of how a perceiver considers all these elements and more. Yet amongst the ecology, a uniform may be the only genuine cue as to whether someone is an employee or not (in Brunswik’s terminology, an *ecologically valid* cue). However, other features such as being alone or being young might bias a perceiver into making an inaccurate judgment. Thus we can test the relationship between participant perceptions (perceived as employee) and the stimulus aspects (uniform, company, age) and the relationship between stimulus aspects and the stimulus qualities (is an employee) which allows us to identify: 1) useful cues (relating to both perceptions and target qualities), 2) irrelevant cues (neither relating to perceptions nor target qualities), 3) missed cues (ecologically valid cues which do not relate to perceptions) and 4) biases (cues which influence

perceptions but are not ecologically valid). This framework allows the study of accuracy and bias in everyday contexts, based on the natural variation between people and stimuli in places.

In the context of first impressions, when asking participants to consider their judgments of new people, we might consider coding many different aspects of our targets and investigate how the different cues from a person might hold more or less attention from our participants. Moreover, in cases where we test judgment ‘accuracy’ (i.e., self-other agreement or task-other agreement) we can evaluate what cues might afford accurate perceptions, missed cues, and biases in interactions. This framework of coding and testing cues outside our primary cues of interest is important as it will identify how relevant our cues of interest are in a noisier, wider environment. Perhaps in everyday experience, the cacophony of other information washes out the explanatory variance of our cues of interest, as participants are biased towards other sources of information. Such methodologies can be standardised enough to be compared across contexts and countries for generalisability (similar spaces, same coding schemes). The coding books and final analysis code can be shared openly and transparently. With participant consent, the recordings of interactions could be stored and used in wider research activities for cross-laboratory coding and reliability checking. Whilst there has been great progress in terms of controlled experimental reliability in the field of first impressions, there are other ways of doing reliability checks on less controlled experiments.

Previously, researchers have used lens models and less controlled (‘zero acquaintance’) interaction settings in their first impressions work, but these examples are the exception rather than the rule in the literature. As Nestler and Back (2013) highlight, deconstructing the lens of first impressions enables us to understand how different types of interpersonal judgments might be more or less accurate by identifying relevant cues. They note how, for example, first impressions of extraversion are more accurate than agreeableness, perhaps because the former traits have more behavioural manifestations identified in lens modelling. Many studies using round robin or dyadic interactions stem from influential work by Albright and colleagues (1988). They noted, almost 35 years ago, much as we do here, that “Using photographs increases experimental control over the information available to the perceiver (i.e., behavioral cues), but may have limited generalizability. Actual physical presence, conversely, decreases control over available information, but is certainly representative of the process of making judgments of strangers” (p. 388). Dynamic studies of first impressions are possible and have been conducted with some impressively large sample sizes—such as Tissera et al. (2021), who studied nearly 8000 dyads’ first interactions—however, such methods are not yet common throughout most of the literature on ‘first impressions’. In general, we advocate for this approach to be adopted more widely in light of trying to improve validity in a field which has made considerable efforts to improve reliability.

Inductive, Participant-Led, Approaches to First Impressions

As a way of developing more inductive approaches to understanding first impressions, some recent work has shifted to participant- or data-led methods. These move from researcher-defined key aspects of a person to study to using participants' responses to guide identification of key aspects of a target person in data. For example, recent work that showed that the traits of symmetry, averageness, and dimorphism—all mostly studied with atomised presentations—showed little predictive capability of attractiveness perceptions (Jones & Jaeger, 2019). Holzleitner and colleagues (2019) used a data-driven approach to explaining variance in facial attractiveness judgments by using principal components analysis to map sources of variability from key marker points (i.e. corners of eyes, tops of lips, sides of nose, etc) across photographs of faces. They uncovered new parameters that explained greater variance in attractiveness than typical deductive aspects of interest mentioned above.

There are more opportunities for being participant-led when collecting judgment data from participants as well. To collect participants' responses to stimulus people, researchers usually give a forced choice or forced spectrum of responses against key anchors of interest, for example, participants are asked to consider how 'attractive' or 'threatening' the stimulus person is. However, it may not be the case that participants' first thoughts on seeing a person are related to these chosen adjectives. As such, research is losing out on key judgments that are important in everyday life. For example, fundamental models of facial first impressions, like the valence-dominance model (Oosterhof & Todorov, 2008), are derived from ratings provided by participants on a series of 14 core traits (e.g., trustworthiness, dominance, unhappy, confident) that are used across many studies. However, C. Lin and colleagues (2021) showed that by using a wider range of trait ratings, significant differences in the key "structure" of impressions emerged, such that four correlated components appeared, indexing traits like youth, age, warmth, and femininity. With advances in text processing techniques, it is also becoming increasingly possible to reproducibly manage data generated from unconstrained responses from participants. This may help develop more theory around an underlying "topic" structure of first impressions that may yield further insights into the underlying psychology.

Moreover, there are opportunities to use more inductive approaches to designing methodological settings and frameworks for first impressions tasks. This would involve rethinking method development by asking focus groups to reflect on what settings, judgments, and dynamics are important for them. This can allow us to move away from a researcher-led approach to designing research that is more data-rich. Whereas a researcher might think judgments of certain adjectives made on observing static faces are important, participants might express particular behaviours or movements as important. Stakeholder- or patient-informed designs are increasingly popular in applied research like forensic and health psychology (see dosReis et al., 2020), and these tools may help further develop first impressions research.

Theory Development and ‘Can’ vs. ‘Do’ Questions

We can broadly conceptualize research as asking two classes of questions—‘Can’ and ‘Do’ questions. This dichotomy is more useful than a general consideration of what is ‘lab’ or ‘field’ work as it focuses on the nature of questions we are trying to answer. ‘Can’ questions ask if it is possible for an effect to occur. *Can* people make first impressions of others in < 100ms (if that is what we provide them with)? *Can* people agree on ratings of attractiveness if only given a voice clip (if that is what we provide them with)? *Can* people use photographs of faces to accurately identify aggression in a target person (if provided only with a face)? These questions increase our knowledge base, but in a specific way. These controlled studies do all they can to maximise the chance of an effect occurring by limiting statistical error (experimental control, large sample size, high target number, etc.). This tells us, all things being equal these effects *can* occur.

In contrast, we can ask ‘do’ questions. *Do* people make first impressions of others in < 100ms (in everyday experience)? *Do* people use voices to decide how attractive a person is (in everyday experience)? *Do* people use faces to accurately detect how aggressive a person is (in everyday experience)? These are questions for which we do not have much empirical evidence. Given that these questions are often those that lay people and practitioners ask of us when we are applying our research—Does this bias our job selection process? Does this affect how my first date goes? Does this lead to a miscarriage of justice?—it should follow that researchers would start by asking ‘do’ questions before leading to ‘can’ questions. Rather than starting from the assumptions of ‘can’ questions and iteratively chipping away at larger wholes, it could improve the theoretical gains of our research to conduct reproducible ‘do’ studies first, to pivot to ‘can’ questions. This matters for sharpening up the resource use of the field. Higher quality, reliable research is more resource intensive. Effective selection of ‘can’ questions through asking broader ‘do’ questions could greatly improve the efficiency of our research questions and theory building.

Conclusion

Much of everyday life is shaped by interactions with other people. Some of the most important parts of a person’s life happen when interacting with unknown others. As psychologists, we want to understand those events better and understand how someone might form a first impression and perceive an interaction partner. This research has been conducted for over 50 years with considerable resources being used to try and understand the phenomenon. But, whilst the credibility crisis in psychology has motivated improvements to reliability in the field, improvements to validity have received less attention and are needed in order to be a more credible science for everyday life.

The normalising of large international collaborative projects, open stimulus databases, and the sharing of materials and code allow for a future of person perception research that can be more readily assessed for its reliability. However, current methodologies still have challenges with stimulus diversity and participants' tasks—we still rarely ask participants to 'mix it up with folks'. This can come at the cost of our theoretical gains. The pursuit of reliability of methods (more than reliability of theoretical outcomes) means that our current reforms have the unintentional consequences of limiting validity. It would be difficult to take these findings on atomised body parts to make a prediction of body wholes for any given first impression. Moreover, we might be over- or under-estimating the relevance of our work for everyday experiences and this can have an impact on how we understand social interactions at large and theories of those who have social difficulties.

We recommend more engagement in dynamic interactive designs, where participant-led approaches allow for codifying rather than controlling the settings where participants might interact. In general, the field should be mindful of what kind of questions a study is designed to address, that is, whether we are asking if an element of a person *can* impact perceptions under certain constraints or whether it *does* have an impact in everyday experience. Given the swift improvements to the field of first impressions in using more reliable methods, we are hopeful that we will see more dynamic, complex paradigms in the future to avoid the unintended consequences of pursuing reliability alone.

Funding: The authors have no funding to report.

Acknowledgments: The authors have no additional (i.e., non-financial) support to report.

Competing Interests: The authors have declared that no competing interests exist.

Author Contributions: Authors after the first author are listed in alphabetical order and author contributions are as follows: structure and position development: LS, primary text drafting: LS, BJ, literature suggestions and text revisions: LS, BJ, AJ, BL, CS. All authors contributed intellectual and written contributions and approved the final text.

References

- Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology*, *55*(3), 387–395.
<https://doi.org/10.1037/0022-3514.55.3.387>

- Alper, S., Bayrak, F., & Yilmaz, O. (2021). All the Dark Triad and some of the Big Five traits are visible in the face. *Personality and Individual Differences*, 168, Article 110350.
<https://doi.org/10.1016/j.paid.2020.110350>
- Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *The Journal of Child Psychology and Psychiatry*, 38(7), 813–822.
<https://doi.org/10.1111/j.1469-7610.1997.tb01599.x>
- Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, 20(1), 1–68. <https://doi.org/10.1177/1529100619832930>
- Batres, C., & Shiramizu, V. (2022). Examining the “attractiveness halo effect” across cultures. *Current Psychology*. Advance online publication. <https://doi.org/10.1007/s12144-022-03575-0>
- Bovet, J., Tognetti, A., & Pollet, T. V. (2022). Methodological issues when using face prototypes: A case study on the Faceaurus dataset. *Evolutionary Human Sciences*, 4, Article e48.
<https://doi.org/10.1017/ehs.2022.25>
- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 62(3), 193–217. <https://doi.org/10.1037/h0047470>
- Carlsson, M., & Eriksson, S. (2019). In-group gender bias in hiring: Real-world evidence. *Economics Letters*, 185, Article 108686. <https://doi.org/10.1016/j.econlet.2019.108686>
- Caton, N. R., Pearson, S. G., & Dixon, B. J. W. (2022). Is facial structure an honest cue to real-world dominance and fighting ability in men? A pre-registered direct replication of. *Evolution and Human Behavior*, 43(4), 314–324. <https://doi.org/10.1016/j.evolhumbehav.2022.04.002>
- Chevallier, C., Kohls, G., Troiani, V., Brodtkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, 16(4), 231–239.
<https://doi.org/10.1016/j.tics.2012.02.007>
- Chita-Tegmark, M. (2016). Social attention in ASD: A review and meta-analysis of eye-tracking studies. *Research in Developmental Disabilities*, 48, 79–93.
<https://doi.org/10.1016/j.ridd.2015.10.011>
- Collin, L., Bindra, J., Raju, M., Gillberg, C., & Minnis, H. (2013). Facial emotion recognition in child psychiatry: A systematic review. *Research in Developmental Disabilities*, 34(5), 1505–1520.
<https://doi.org/10.1016/j.ridd.2013.01.008>
- Cook, R., & Over, H. (2021). Why is the literature on first impressions so focused on White faces? *Royal Society Open Science*, 8(9), Article 211146. <https://doi.org/10.1098/rsos.211146>
- Cowen, A. S., Keltner, D., Schroff, F., Jou, B., Adam, H., & Prasad, G. (2021). Sixteen facial expressions occur in similar contexts worldwide. *Nature*, 589(7841), 251–257.
<https://doi.org/10.1038/s41586-020-3037-7>
- Dawel, A., Miller, E. J., Horsburgh, A., & Ford, P. (2022). A systematic survey of face stimuli used in psychological research 2000–2020. *Behavior Research Methods*, 54(4), 1889–1901.
<https://doi.org/10.3758/s13428-021-01705-3>

- dosReis, S., Butler, B., Caicedo, J., Kennedy, A., Hong, Y. D., Zhang, C., & Slejko, J. F. (2020). Stakeholder-engaged derivation of patient-informed value elements. *The Patient - Patient-Centered Outcomes Research*, 13(5), 611–621. <https://doi.org/10.1007/s40271-020-00433-8>
- Durán, J. I., & Fernández-Dols, J.-M. (2021). Do emotions result in their predicted facial expressions? A meta-analysis of studies on the co-occurrence of expression and emotion. *Emotion*, 21(7), 1550–1569. <https://doi.org/10.1037/emo0001015>
- Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management*, 55, 62–73. <https://doi.org/10.1016/j.tourman.2016.01.013>
- Foo, Y. Z., Sutherland, C. A. M., Burton, N. S., Nakagawa, S., & Rhodes, G. (2022). Accuracy in facial trustworthiness impressions: Kernel of truth or modern physiognomy? A meta-analysis. *Personality and Social Psychology Bulletin*, 48(11), 1580–1596. <https://doi.org/10.1177/01461672211048110>
- Gernsbacher, M. A., & Yergeau, M. (2019). Empirical failures of the claim that autistic people lack a theory of mind. *Archives of Scientific Psychology*, 7(1), 102–118. <https://doi.org/10.1037/arc0000067>
- Hammond, K. R. (1948). Subject and object sampling—A note. *Psychological Bulletin*, 45(6), 530–533. <https://doi.org/10.1037/h0056803>
- Holzleitner, I. J., Lee, A. J., Hahn, A. C., Kandrik, M., Bovet, J., Renoult, J. P., Simmons, D., Garrod, O., DeBruine, L. M., & Jones, B. C. (2019). Comparing theory-driven and data-driven attractiveness models using images of real women's faces. *Journal of Experimental Psychology: Human Perception and Performance*, 45(12), 1589–1595. <https://doi.org/10.1037/xhp0000685>
- Hutchins, T. L., & Brien, A. (2016). Conversational topic moderates social attention in autism spectrum disorder: Talking about emotions is like driving in a snowstorm. *Research in Autism Spectrum Disorders*, 26, 99–110. <https://doi.org/10.1016/j.rasd.2016.03.006>
- Jaeger, B. (2020). *Trait ratings for the radboud faces database*. PsyArXiv. <https://psyarxiv.com/cf5ad/>
- Jaeger, B., & Jones, A. L. (2022). Which facial features are central in impression formation? *Social Psychological and Personality Science*, 13(2), 553–561. <https://doi.org/10.1177/19485506211034979>
- Jaeger, B., Slegers, W. W. A., Evans, A. M., Stel, M., & van Beest, I. (2019). The effects of facial attractiveness and trustworthiness in online peer-to-peer markets. *Journal of Economic Psychology*, 75, Article 102125. <https://doi.org/10.1016/j.joep.2018.11.004>
- Jones, A., & Jaeger, B. (2019). Biological bases of beauty revisited: The effect of symmetry, averageness, and sexual dimorphism on female facial attractiveness. *Symmetry*, 11(2), Article 279. <https://doi.org/10.3390/sym11020279>
- Jones, B. C., DeBruine, L. M., Flake, J. K., Liuzza, M. T., Antfolk, J., Arinze, N. C., Ndukaihe, I. L. G., Bloxson, N. G., Lewis, S. C., Feroni, F., Willis, M. L., Cubillas, C. P., Vadillo, M. A., Turiegano, E., Gilead, M., Simchon, A., Saribay, S. A., Owsley, N. C., Jang, C., ...Coles, N. A. (2021). To which world regions does the valence–dominance model of social perception apply? *Nature Human Behaviour*, 5(1), 159–169. <https://doi.org/10.1038/s41562-020-01007-2>

- Kikuchi, Y., Akechi, H., Senju, A., Tojo, Y., Osanai, H., Saito, A., & Hasegawa, T. (2022). Attention to live eye contact in adolescents with autism spectrum disorder. *Autism Research, 15*(4), 702–711. <https://doi.org/10.1002/aur.2676>
- Kramer, R. S. S., & Gardner, E. M. (2020). Facial trustworthiness and criminal sentencing: A comment on Wilson and Rule (2015). *Psychological Reports, 123*(5), 1854–1868. <https://doi.org/10.1177/0033294119889582>
- Kumle, L., Vö, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods, 53*(6), 2528–2543. <https://doi.org/10.3758/s13428-021-01546-0>
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion, 24*(8), 1377–1388. <https://doi.org/10.1080/02699930903485076>
- Lin, C., Keles, U., & Adolphs, R. (2021). Four dimensions characterize attributions from faces using a representative set of English trait words. *Nature Communications, 12*, Article 5168. <https://doi.org/10.1038/s41467-021-25500-y>
- Lin, H., Werner, K. M., & Inzlicht, M. (2021). Promises and perils of experimentation: The mutual-internal-validity problem. *Perspectives on Psychological Science, 16*(4), 854–863. <https://doi.org/10.1177/1745691620974773>
- 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*(4), 1122–1135. <https://doi.org/10.3758/s13428-014-0532-5>
- Marsden, J., Glazebrook, C., Tully, R., & Völlm, B. (2019). Do adult males with antisocial personality disorder (with and without co-morbid psychopathy) have deficits in emotion processing and empathy? A systematic review. *Aggression and Violent Behavior, 48*, 197–217. <https://doi.org/10.1016/j.avb.2019.08.009>
- Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., Grahe, J. E., McCarthy, R. J., Musser, E. D., Antfolk, J., Castille, C. M., Evans, T. R., Fiedler, S., Flake, J. K., Forero, D. A., Janssen, S. M. J., Keene, J. R., Protzko, J., Aczel, B., ...Chartier, C. R. (2018). The psychological science accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science, 1*(4), 501–515. <https://doi.org/10.1177/2515245918797607>
- Neisser, U. (1980). On ‘social knowing’. *Personality and Social Psychology Bulletin, 6*(4), 601–605. <https://doi.org/10.1177/014616728064012>
- Nestler, S., & Back, M. D. (2013). Applications and extensions of the lens model to understand interpersonal judgments at zero acquaintance. *Current Directions in Psychological Science, 22*(5), 374–379. <https://doi.org/10.1177/0963721413486148>
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*(32), 11087–11092. <https://doi.org/10.1073/pnas.0805664105>
- Plana, I., Lavoie, M.-A., Battaglia, M., & Achim, A. M. (2014). A meta-analysis and scoping review of social cognition performance in social phobia, posttraumatic stress disorder and other anxiety

- disorders. *Journal of Anxiety Disorders*, 28(2), 169–177.
<https://doi.org/10.1016/j.janxdis.2013.09.005>
- Satchell, L. (2019). From photograph to face-to-face: Brief interactions change person and personality judgments. *Journal of Experimental Social Psychology*, 82, 266–276.
<https://doi.org/10.1016/j.jesp.2019.02.010>
- Singh, B., Gambrell, A., & Correll, J. (2022). Face templates for the Chicago Face Database. *Behavior Research Methods*. Advance online publication. <https://doi.org/10.3758/s13428-022-01830-7>
- Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science*, 21(3), 349–354. <https://doi.org/10.1177/0956797610362647>
- Sutherland, C. A. M., Liu, X., Zhang, L., Chu, Y., Oldmeadow, J. A., & Young, A. W. (2018). Facial first impressions across culture: Data-driven modeling of Chinese and British perceivers' unconstrained facial impressions. *Personality and Social Psychology Bulletin*, 44(4), 521–537.
<https://doi.org/10.1177/0146167217744194>
- Tissera, H., Gazzard Kerr, L., Carlson, E. N., & Human, L. J. (2021). Social anxiety and liking: Towards understanding the role of metaperceptions in first impressions. *Journal of Personality and Social Psychology*, 121(4), 948–968. <https://doi.org/10.1037/pspp0000363>
- Varela, V. P. L., Towler, A., Kemp, R. I., & White, D. (2023). Looking at faces in the wild. *Scientific Reports*, 13(1), Article 783. <https://doi.org/10.1038/s41598-022-25268-1>
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, 17(7), 592–598. <https://doi.org/10.1111/j.1467-9280.2006.01750.x>
- Zäske, R., Skuk, V. G., Golle, J., & Schweinberger, S. R. (2020). The Jena Speaker Set (JESS)—A database of voice stimuli from unfamiliar young and old adult speakers. *Behavior Research Methods*, 52(3), 990–1007. <https://doi.org/10.3758/s13428-019-01296-0>

PSPS
 Polish
 Social
 Psychological
 Society

Social Psychological Bulletin (SPB)
 is an official journal of the Polish
 Social Psychological Society
 (PSPS).



leibniz-psychology.org

PsychOpen GOLD is a publishing
 service by Leibniz Institute for
 Psychology (ZPID), Germany.