2 From Spontaneous Trait Inferences to Spontaneous Person Impressions

Alexander Todorov

The University of Chicago Booth School of Business

I have great memories from my graduate student days at New York University. Our student offices didn't have windows and I could clearly hear the answering machines in the adjacent offices, but we were happy and productive. The social psychology area was highly collaborative and most of us worked with multiple faculty members. I worked with John Bargh, Shelly Chaiken, Yaacov Trope, and Jim Uleman. The lab groups were collaborative, with some weekly meetings being highly argumentative, while others calm. Regardless, they all were fun.

At the time, I was fascinated by subliminal priming effects and models of assimilation and contrast effects in judgments. However, my fascination met with mixed empirical success. Subliminal priming effects were generally weak and that made design of complex experiments with multiple factors challenging. In response to these normal quirks of experimental exploration, I started what was as a side project at the time. I started a series of studies with Jim Uleman on spontaneous trait inferences. This side project turned into my dissertation and shaped my research for the next 20 years.

Spontaneous Person Inferences

Jim was a pioneer in research on unintentional higher-level inferences (Uleman et al., 1996; Winter & Uleman, 1984), and we set out to work on an unresolved question in the area of spontaneous trait inferences. We asked how people form trait inferences (e.g., "honest") from behavioral statements (e.g., "Bob returned the lost wallet."). It was well established that such inferences are made spontaneously upon reading a trait-implying behavioral statement, but it was not clear whether such spontaneous inferences are free-floating inferences that are simply temporary accessible in memory or are bound to the representation of the agent who performed the behavior. This question was important, because the two possibilities have radically different implications for person perception in particular and social cognition in general. According to the first possibility of free-floating inferences, spontaneous inferences are important in the immediate situation but have no long-term implications for person representations. According to the second

DOI: 10.4324/9781003045687-3

possibility of agent-bound representations, these inferences modify person representations. The evidence from cued recall paradigms in which the inference ("honest") cues the recall of the agent and the behavior was mixed with respect to these two possibilities (Winter & Uleman, 1984). Whereas this inference did facilitate the recall of the behavior, it did not seem to facilitate the recall of the agent. Or at least it only seemed to do so in perceivers with motives that encouraged the encoding of links to the agent (Moskowitz, 1993).

To test whether trait inferences are bound to the representation of the agents, we designed a false recognition paradigm. In this paradigm, participants study faces with trait-diagnostic behaviors for a subsequent memory task. Later in a recognition test, they see face-trait pairs (e.g., Bob's photo and "honest") and decide whether they saw the trait in the sentence presented with Bob's face. In our first paper (Todorov & Uleman, 2002), we showed that participants were more likely to falsely recognize implied traits when presented with the agent's photo than when presented with other familiar photos. This effect was robust and large (the average effect size across experiments in terms of Pearson's r was 0.66)—an effect size in stark contrast to what I was observing with my attempts to create subliminal priming paradigms. The effect did not seem to depend on the number of faces and behaviors participants were exposed to. Across experiments, this number varied from 36 to 120. Moreover, the effect was not dependent on explicit memory for the behaviors. Even when participants did not recall or recognize the specific behavior, they were more likely to associate the implied trait with the agent's face. Analyses at the level of the stimuli (behavioral statements) showed that false recognition rates of implied traits were predicted by the strength of the trait implications of the behavioral statements (as measured by explicit judgments of a separate group of participants), showing that spontaneous inferences are highly specific and their strength varies as a function of the behavioral evidence. These findings, coupled with findings from the savings in relearning paradigm (Carlston & Skowronski, 1994; Carlston et al., 1995), demonstrated that spontaneous trait inferences are bound to the representation of the person enacting the behavior.

Encouraged by the robustness of the evidence for links between inferred traits and agents' faces, in our second paper (Todorov & Uleman, 2003), we studied to what extent the processes leading to these links are relatively independent of attentional resources. In earlier experiments, we presented the faces and behaviors for 5 or 10 seconds (if self-paced, participants typically spend a little over 6 seconds per face and behavior). In our first experiment (Todorov & Uleman, 2003), we included a condition, in which each face-behavior pair was presented for only 2 seconds. Nonetheless, participants were more likely to falsely recognize implied traits in the context of the agent's face than in the context of another familiar face. In our second experiment, we induced shallow processing of the information by asking participants to count the number of nouns in each sentence. Although this manipulation reduced the false recognition effect, it did not eliminate it. In the third experiment, we introduced cognitive load. Participants were asked to rehearse six-digit numbers while reading the behavioral statements. Once again, the false recognition effect was present and large in size. In a final experiment, we collected person and behavior judgments of the behavioral statements. When asked to make a judgment about a person from the statement "Bob returned the lost wallet," participants considered the question, "Is Bob an honest person?" In contrast, when asked to make a judgment about a behavior, they considered the question, "Is this an honest behavior?" We used these two types of judgments to predict false recognition rates across experiments, including our initial experiments in Todorov and Uleman (2002). Person judgments, but not behavior judgments, predicted the false recognition rates, showing that people infer and associate traits with agents' faces rather than simply associate the meaning of behaviors with faces. These findings clearly supported the hypothesis that spontaneous trait inferences modify specific person representations.

Yet it was not clear from our previous studies whether trait associations are specifically bound to the representation of the face of the agent who performed the behavior rather than to any face that happened to be co-present with the behavior. In our final paper (Todorov & Uleman, 2004), we modified the learning trials of the false recognition paradigm to include two faces and a behavior referring to one of the faces. Participants were more likely to associate the traits with the face of the person who performed the behavior than with the control face. This effect, though reduced, persisted after a week. Interestingly, after a week the hit rate of correct recognition of presented traits was indistinguishable from the false recognition rate of implied traits. But in both cases, these traits were more likely to be associated with the right face. We also ruled out that our findings could be explained by differential attention to the faces. In the final two experiments, on each learning trial participants were presented with two faces and two behaviors, each referring to one of the faces. This paradigm forced participants to pay attention to both of the faces and behaviors. Nonetheless, we again found that participants were more likely to associate the implied traits with the faces of the actors who performed the trait-implying behaviors. Finally, we obtained the same results when we used different images of the same face identity during learning and testing, showing that spontaneously inferred traits are associated with abstract person representations rather than with specific image representations of faces.

Findings from the two most prominent paradigms for detecting spontaneous trait inferences—false recognition (Todorov & Uleman, 2002) and savings in relearning (Carlston & Skowronski, 1994)—clearly demonstrate that such inferences are bound to the representations of the agents who enacted the behavior. There are differences between these paradigms (see Crawford et al., 2007; Goren & Todorov, 2009), but the similarities are more important. Both rely on the retrieval of traits implied by behaviors, and these

traits are cued by photos of the agents initially presented with the behaviors. The key to success of both of these paradigms is not so much the specific measures they use, but the presence of faces. In contrast to names or other labels such as occupations, faces are highly distinctive, memorable, and the natural stimuli around which to organize person memories.

The Importance of Faces

The work with Jim led me to conduct systematic studies of the importance of the face in social cognition (Todorov, 2017; Todorov et al., 2015). This work, as well as the training in John Bargh's and Yaacov Trope's labs, also introduced persistent themes in my research: the efficiency and the importance of social judgments.

In experiments conducted with Jim, we were not interested in facial appearance per se. Typically, we randomly assigned behaviors to faces, as well as counterbalanced faces and behaviors, to make sure that the observed effects are due to the behaviors paired with the faces. But faces are a rich source of social inferences. Already in the 1950s, Paul Secord conducted a number of studies demonstrating that people infer traits from faces (e.g., Secord, 1958). Leslie Zebrowitz conducted seminal studies in the 1980s showing how specific facial characteristics such as baby-faced features trigger specific trait inferences such as "naïve" (e.g., Berry & Zebrowitz McArthur, 1985, 1986; Montepare & Zebrowitz McArthur, 1986; Zebrowitz McArthur & Apatow, 1984).

Following in their steps, I started systematic studies on inferences from faces in my newly formed lab at Princeton Univesity (Oosterhof & Todorov, 2008; Todorov et al., 2008). Two sets of initial findings demonstrated the importance and efficiency of social judgments from faces. In the first set of findings, we showed that naïve judgments of competence based solely on the facial appearance of politicians predicted electoral success (Todorov et al., 2005). The findings were surprising, but replicated in many different contexts (Antonakis & Dalgas, 2009; Lawson et al., 2010; Olivola & Todorov, 2010a; Poutvaara et al., 2009; Sussman et al., 2013). Studies by political scientists showed that the effects of appearance on voting decisions are limited to those voters who know next to nothing about politics and are exposed to images of the politicians (Ahler et al., 2017; Lenz & Lawson, 2011), a great example of heuristic processing where shallow, rapid inferences substitute more cognitively demanding inferences from substantive information (Hall et al., 2009).

The second set of findings was that people need minimal exposure to faces to form specific trait inferences such as trustworthiness (Willis & Todorov, 2006). In our initial studies, we presented faces for 100, 500, or 1,000 ms. Contrary to our expectations, judgments did not differ as a function of the length of exposure. The only effect of the latter was to increase confidence in judgments. Subsequent studies used better masking procedures and presented faces for even shorter exposures (Ballew & Todorov, 2007; Bar et al., 2006; Borkenau et al., 2009; Porter et al., 2008; Rule et al., 2009; Todorov et al., 2010; Todorov et al., 2009).

Generally, as little as 34 ms exposure is sufficient for people to form a judgment that is correlated with judgments made in the absence of time constraints, and this correlation doesn't increase in magnitude with exposures longer than about 200 ms (Todorov et al., 2009; 2010). Trait inferences from faces are literally single glance impressions.

Although there is little evidence that trait inferences from facial appearance are accurate (Hassin & Trope, 2000; Olivola & Todorov, 2010b; Todorov, 2017; Todorov et al., 2015), these initial findings showed that these inferences are highly efficient and matter for important social outcomes. In terms of the construction of social judgments, the findings also showed that people agree on these judgments. This agreement formed the basis of one of the questions that has guided much of the research in my lab for more than a decade (Oosterhof & Todorov, 2008; Todorov & Oh, 2021). The question was, given the agreement in judgments, how can we identify the perceptual basis or the configurations of facial features that lead to specific trait inferences.

To answer this question, we developed data-driven computational methods, which do not depend on prior hunches of what facial features are important for judgments (Oosterhof & Todorov, 2008; Todorov et al., 2011; Todorov & Oh, 2021; Todorov & Oosterhof, 2011). These methods were necessary, because it was practically impossible to discover the configurations of features that matter for judgments in the standard hypothesis-driven framework. In the latter framework, one posits that a set of features (e.g., shape of mouth, shape of eyebrows) influences judgments (e.g., friendliness) and then manipulates these features to test their effects on judgments. But manipulating just 10 binary facial features in a factorial design results in over 1,000 combinations; and manipulating 20 binary features results in over a million. Moreover, features are not binary and we don't even know what constitutes a feature (e.g., mouth vs. lips vs. corner of lips). Finally, features would not even be manipulated, if the experimenter doesn't think that they are important for judgments.

In our data-driven framework, we used a statistical model of face representation, in which each face is represented as a 100-dimensional vector. The appearance of each face is perfectly determined by its coordinates in this multi-dimensional face space. Rather than manipulating features, we simply randomly sampled faces from the multi-dimensional face space, and asked participants to judge the faces on various trait dimensions. Given the average trait judgment, we can then build a model of this judgment that captures the variation in appearance that is important for the judgment. The process is akin to finding the regression line, predicted from 100 orthogonal predictors (the coordinates of the faces), that accounts for most variance in judgments (for a detailed review of the methods, see Todorov & Oh, 2021).

Over the years, we have generated dozens of models of trait judgments (Funk et al., 2016; Oh, Buck, et al., 2019; Oh, Dotsh, et al., 2019; Said & Todorov, 2011; Todorov et al., 2013). These models can manipulate the appearance of novel faces parametrically, increasing or decreasing their perceived value on trait dimensions such as trustworthiness and competence. Based on the models, we have created many databases of faces parametrically manipulated on trait dimensions and made those available for academic research. More than 4,000 users from over 900 institutions have used these databases for research, addressing a variety of questions: from studying infants' sensitivity to facial signals of trustworthiness and dominance (Jessen & Grossmann, 2016) to the effects of appearance on economic decisions (Rezlescu et al., 2012) and voting preferences (Laustsen & Petersen, 2016).

Although the models described previously are models of explicit judgments, they are easily extendable to implicit measures of judgments. Moreover, models of implicit measures could be immediately related to models of explicit judgments, because both are in the same statistical multidimensional space. That is, the similarity of models (whether based on explicit or implicit measures) is immediately given; it is simply captured by the correlation of the models (e.g., each model is a vector in the same multidimensional face space).

In recent work, harking back to my days at New York University, Ran Hassin and I collaborated to build a model of faces that break faster into consciousness (Abir et al., 2018). At New York University, Ran and I spent a lot of time arguing with each other; and unconscious processes were a core interest of the social cognition group back then (Hassin et al., 2005). Using continuous flash suppression, which suppresses visual input from one of the eyes, we measured the response times to detecting faces breaking into consciousness (e.g., being seen by the suppressed eye). We built a model of these response times, capturing the variation in facial appearance that emerges faster in consciousness. This model was highly correlated with a model of judgments of dominance. Recently, we have also built models of neural measures to faces (Cao et al., 2020). Such models of implicit measures can capture the content of truly spontaneous impressions.

My work with Jim was about associating trait inferences with person representations. As it turned out, faces were the critical stimuli to detect these associations. The importance of faces led me to studying trait inferences based solely on facial appearance, but I never abandoned the original question we studied with Jim.

The Robustness of Associating Affective Inferences with Faces

The models of judgments from faces are extremely powerful, but they also mask individual differences in trait inferences from facial appearance, simply because these are models of aggregated judgments. The typical measure of agreement in judgments is Cronbach's alpha. A high alpha of 0.90 simply indicates the expected correlation between the aggregated judgments of two groups (with the same size) of raters. But the average correlation between raters within a group would be much smaller, typically of the order 0.30. In fact, partitioning the reliable variance in judgments from faces to shared variance with others and to idiosyncratic (individually stable) variance shows that the only judgment, in which these are relatively equal, is attractiveness. For any other social judgment, such as approachability, the idiosyncratic variance is much larger than the shared variance (Martinez et al., 2020).

What determines idiosyncratic contributions to trait inferences from facial appearance? One possibility is similarity to the faces of significant others, another theme that has its origins at New York University (Andersen & Cole, 1990; Chen & Andersen, 1999; Kraus & Chen, 2010). To the extent that different individuals have different-looking significant others, friends, and foes and these individuals use the similarity of strangers to their familiar others to make trait inferences based on this resemblance, there should be systematic individual differences in trait inferences. To experimentally test this hypothesis, we followed the logic of our studies with Jim on inducing trait inferences from behavioral statements (Verosky & Todorov, 2010). In the first stage of our experiments, we had participants associate faces with positive, negative, or neutral behavioral statements. Then, we asked them to make judgments of novel faces, which were subtly morphed with the familiar faces. Participants judged novel faces more positively when they were morphed with faces associated with positive information and more negatively when they were morphed with faces associated with negative information. In a subsequent study, we showed that this learning generalization from familiar others occurred even when participants were explicitly asked to disregard facial similarity information and made their judgments under cognitive load (Verosky & Todorov, 2013). Such processes of learning generalization based on similarity to familiar others are one of the mechanisms underlying learning to trust (FeldmanHall et al., 2018).

The studies described previously led me to a series of studies, which are a direct descendant of my work with Jim. In these studies (Falvello et al., 2015; Ferrari et al., 2020; Verosky et al., 2018), we did not use a false recognition paradigm, but we studied highly related questions about the nature of the associations between faces and the evaluative trait implications of behaviors. In the experiments, participants were first presented with faces and behaviors, which varied in valence, and then evaluated the faces without the presence of the behaviors.

As described in the previous section, people need minimal exposure to faces to form trait inferences. The mechanisms underlying this finding are straightforward to explain, given the computational work on models of judgments. The trait inferences are triggered by specific configurations of facial features. But in the case of trait associations with faces, there is nothing in the physical appearance of the face that "codes" the association. For the association to be retrieved, one needs to access a specific representation of the person who performed the behavior, perhaps requiring extra cognitive resources. To explore this question, we contrasted the effects of inferences from facial appearance and the effects of inferences from behavioral information (Verosky et al., 2018). Rather surprisingly, the effect of inferences from behaviors was detectable after 35 ms exposure to the face: participants evaluated more positively faces associated with positive behaviors than faces associated with negative behaviors. If anything, this effect was larger than the effect of appearance (evaluating "trustworthy-looking" faces more positively than "untrustworthy-looking"). In a second study, we introduced a response deadline procedure forcing participants to make rapid judgments. Nonetheless, the effect of inferences from behaviors was detectable after 35 ms exposure to the face, although the effect was reduced in size. Finally, we measured the recognition of the faces. Not surprisingly, as face recognition increased, so did the effect of inferences from behaviors: the difference between the evaluation of faces associated with positive behaviors and the evaluation of faces associated with negative information increased. But the effect of inferences from behaviors was detectable at exceedingly low levels of face recognition. This effect emerged when participants reported recognizing the faces at a recognition value of three (on a nine-point scale), which was below the average value of recognition for novel faces. This was also the case for face exposures as short as 27 ms. These findings show how powerful social learning is in modifying person representations.

In our studies with Jim on spontaneous trait inferences, we used as many as 120 face-behavior pairs and observed that the effect size of the false recognition effect did not seem to vary as a function of the number of face-behavior pairs. To test whether there are limits on the ability to form affective associations with faces, we presented participants with as many as 500 face and behavior pairs (Falvello et al., 2015). We expected that as the number of faces and behaviors increases, the effect of inferences from behaviors on evaluation of faces would decrease. Surprisingly, we found that this effect was as strong after seeing 400 faces and behaviors as after seeing 100. A post-hoc analysis across three experiments suggested that the effect might start decreasing after seeing 300 faces and behaviors. But given the post-hoc nature of the analysis, it remains to be determined when affective associations with faces start breaking down.

Another surprising finding of the previous study was that we found similar effects for scenes. Participants were able to form affective associations with scenes paired with positive or negative descriptions, and the strength of the effect was similar to the effect for faces. This finding suggested that both kinds of affective associations (with faces and scenes) are driven by the same affect-based mechanisms and that perhaps the rich person-attribution processes, which we posited in the case of spontaneous trait inferences, are not necessary. To test this possibility, we contrasted learning of associations with faces and learning of associations with places (e.g., scenes and houses) (Ferrari et al., 2020). The key manipulation was whether the statements were relevant (e.g., a behavior paired with a face; a positive scene description with a scene) or irrelevant (e.g., a behavior paired with a scene). We found that when statements were repeated, participants formed affective associations with places irrespective of the relevance of the source of these affective associations. This finding is consistent with a simple associative affect-based mechanism. In contrast, affective associations with faces were much stronger when the source of associations was relevant (e.g., behaviors).

Taken together, our findings show that people are remarkably good at forming affective associations with faces from relevant behavioral information, that these associations are specific to the person who performed the behavior, and that they are rapidly triggered by the mere presence of the person's face. All these findings were foreshadowed by my early work with Jim on spontaneous trait inferences and find a new expression in the recent research of Melissa Ferguson, a peer from NYU and a member of the lab groups of Bargh and Trope. Her recent work shows that implicit impressions can be rapidly updated (just like explicit impressions) in light of relevant behavioral information (Ferguson et al., 2019; Shen et al., 2020).

Beyond Inferences from Faces and Behaviors

All of the inferences described previously had to do either with inferences from facial appearance or behavioral statements, but they need not be limited to these two sources of information. People would use whatever information is available to rapidly form coherent person impressions. Two recent research examples are on inferences from bodily information and clothing cues.

Indeed, bodily information informs inferences of emotional expressions (Aviezer et al., 2012a, 2012b; Aviezer et al., 2015; Hassin et al., 2013). The driving force behind these studies was Hillel Aviezer, who was a post-doc with me and Yaacov Trope. Before joining our labs, Aviezer had already shown that people cannot ignore bodily information when inferring facial expressions of emotions (Aviezer et al., 2011; Aviezer et al., 2008). An expression of disgust is instantaneously perceived as anger, if the face expressing disgust is perched on a body about to hit someone. We studied extreme real-life emotions (e.g., winning, losing, pain, pleasure) and found that when people were only shown faces, they could not discriminate between positive and negative emotions (Aviezer et al., 2012a). In contrast, when shown bodies, they were pretty good at discriminating the valence of emotions. Yet, when asked what is the main source of their emotion inferences, the majority of participants believed that it was the face rather than the body. When provided with the intact images (e.g., faces and bodies), participants rapidly disambiguated the emotional expressions without ever occurring to them that the expressions were ambiguous.

The second example is about inferences of competence from clothing cues indicating economic status (Oh et al., 2020). In this work, we asked participants to make judgments of competence from faces. The critical manipulation was that the faces were presented with upper body clothing that was either perceived as "richer" or "poorer," though none of the clothing indicated poverty. The same face was evaluated as more competent when paired with "richer" than with "poorer" clothing. Moreover, in nine experiments, we failed to eliminate this effect. We presented the faces for brief time, we told participants to ignore the clothing, we told them that the people depicted in the photos worked similar jobs and earned similar salaries, and we told them that the clothing was completely undiagnostic for real competence. In one study, we introduced large incentives (the participant who made judgments most similar to judgments of the faces alone was paid $100). None of these manipulations eradicated the effect of clothing on inferences of competence.

Conclusion

Inferences about people are powerful and many of them have the characteristics of automatic processes (Bargh, 1994): they are efficient, often unintentional, often uncontrollable, and often we are not aware of the cues that really influence our judgments. When encountering other people, we grab on whatever information is available at the moment to rapidly form spontaneous person impressions. Spontaneous trait inferences are part of this process. They are not just trait inferences; they are trait inferences that become integrated into the representation of the person. This is precisely their functional significance. After all, information about past actions is a more reliable source of person inferences than facial appearance or clothing.

References

Abir, Y., Sklar, A. Y., Dotsch, R., Todorov, A., & Hassin, R. R. (2018). The determinants of consciousness of human faces. *Nature Human Behaviour*, *2*, 194–199.

Andersen, S. M., & Cole, S. W. (1990). "Do I know you?": The role of significant others in general social perception. *Journal of Personality and Social Psychology*, *59*, 384–399.

Antonakis, J., & Dalgas, O. (2009). Predicting elections: Child's play! *Science*, *323*, 1183–1183.

Ahler, D. J., Citrin, J., Dougal, M. C., & Lenz, G. S. (2017). Face value? Experimental evidence that candidate appearance influences electoral choice. *Political Behavior*, *39*, 77–102.

Aviezer, H., Bentin, S., Dudarev, V., & Hassin, R. R. (2011). The automaticity of emotional face-context integration. *Emotion*, *11*, 1406–1414.

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., Moscowitch, M., & Bentin, S. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological Science*, *19*, 724–732.

Aviezer, H., Messinger, D. S., Zangvil, S., Mattson, W. I., Gangi, D. N., & Todorov, A. (2015). Thrill of victory or agony of defeat? Perceivers fail to utilize information in facial movements. *Emotion*, *15*, 791–797.

Aviezer, H., Trope, Y., & Todorov, A. (2012a). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, 338, 1225–1229.

Aviezer, H., Trope, Y., & Todorov, A. (2012b). Holistic person processing: Faces with bodies tell the whole story. *Journal of Personality and Social Psychology*, *103*, 20–37.

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences, 104*, 17948–17953.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6*, 269–278.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In J. R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., Vol. 1, pp. 1–40). Hillsdale, NJ: Erlbaum.

Berry, D. S., & Zebrowitz McArthur, L. (1985). Some components and consequences of a babyface. *Journal of Personality and Social Psychology, 48*, 312–323.

Berry, D. S., & Zebrowitz McArthur, L. (1986). Perceiving character in faces: The impact of age-related craniofacial changes on social perception. *Psychological Bulletin, 100*, 3–18.

Borkenau, P., Brecke, S., Möttig, C., & Paelecke, P. (2009). Extraversion is accurately perceived after a 50-ms exposure to a face. *Journal of Research in Personality, 43*, 703–706.

Cao, R., Li, X., Todorov, A., & Wang, S. (2020). A flexible neural representation of faces in the human brain. *Cerebral Cortex Communications, 1*, 1–12.

Carlston, D. E., & Skowronski, J. J. (1994). Saving in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology, 66*, 840–856.

Carlston, D. E., Skowronski, J. J., & Sparks, C. (1995). Savings in relearning: II. On the formation of behavior-based trait associations and inferences. *Journal of Personality and Social Psychology, 69*, 420–436.

Chen, S., & Andersen, S. M. (1999). Relationships from the past in the present: Significant-other representations and transference in inter- personal life. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 31, pp. 123–190). San Diego, CA: Academic Press.

Crawford, M. T., Skowronski, J. J., & Stiff, C. (2007). Limiting the spread of spontaneous trait transference. *Journal of Experimental Social Psychology, 43*, 466–472.

Crawford, M. T., Skowronski, J. J., Stiff, C., & Leonards, U. (2008). Seeing, but not thinking: Limiting the spread of spontaneous trait transference II. *Journal of Experimental Social Psychology, 44*, 840–847.

Falvello, V., Vinson, M., Ferrari, C., & Todorov, A. (2015). The robustness of learning about the trustworthiness of other people. *Social Cognition, 33*, 368–386.

FeldmanHall, O., Dunsmoor, J. E., Tompary, A., Hunter, L. E., Todorov, A. T., & Phelps, E. A. (2018). Stimulus generalization as a mechanism for learning to trust. *Proceedings of the National Academy of Sciences, 115*, E1690–E1697.

Ferguson, M. J., Mann, T. C., Cone, J., & Shen, X. (2019). When and how implicit impressions can be updated. *Current Directions in Psychological Science, 28*, 331–336.

Ferrari, C., Oh, D., Labbree, B., & Todorov, A. (2020). Learning the affective value of people: More than affect-based mechanisms. *Acta Psychologica, 203*, 103011.

Funk, F., Walker, M., & Todorov, A. (2016). Modelling perceptions of criminality and remorse from faces using a data-driven computational approach. *Cognition and Emotion, 31*, 1431–1443.

Goren, A., & Todorov, A. (2009). Two faces are better than one: Eliminating false trait associations with faces. *Social Cognition, 27*, 222–248.

Hall, C. C., Goren, A., Chaiken, S., & Todorov, A. (2009). Shallow cues with deep effects: Trait judgments from faces and voting decisions. In E. Borgida, J. L. Sullivan, & C. M. Federico (Eds.), *The political psychology of democratic citizenship* (pp. 73–99). Oxford University Press.

Hassin, R. R., Aviezer, H., & Bentin, S. (2013). Inherently ambiguous: Facial expressions of emotion in context. *Emotion Review, 5*, 60–65.

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology, 78*, 837–852.

Hassin, R., Uleman, J. S., & Bargh, J. A. (Eds.) (2005). *The New Unconscious*. New York: Oxford University Press.

Jessen, S., & Grossmann, T. (2016). Neural and behavioral evidence for infants' sensitivity to the trustworthiness of faces. *Journal of Cognitive Neuroscience, 28*, 1728–1736.

Kraus, M. W., & Chen, S. (2010). Facial-feature resemblance elicits the transference effect. *Psychological Science, 21*, 518–522.

Laustsen, L., & Petersen, M. B. (2016). Winning faces vary by ideology: How nonverbal source cues influence election and communication success in politics. *Political Communication, 33*, 188–211.

Lawson, C., Lenz, G. S., Baker, A., & Myers, M. (2010). Looking like a winner: Candidate appearance and electoral success in new democracies. *World Politics, 62*, 561–593.

Lenz, G. S., & Lawson, C. (2011). Looking the part: Television leads less informed citizens to vote based on candidates' appearance. *American Journal of Political Science, 55*, 574–589.

Martinez, J. E., Funk, F., & Todorov, A. (2020). Quantifying idiosyncratic and shared contributions to judgment. *Behavior Research Methods, 52*, 1428–1444.

Montepare, J. M., & Zebrowitz McArthur, L. (1986). The influence of facial characteristics on children's age perceptions. *Journal of Experimental Child Psychology, 42*, 303–314.

Moskowitz, G. B. (1993). Individual differences in social categorization: The influence of personal need for structure on spontaneous trait inferences. *Journal of Perosnality and Social Psychology, 65*, 132–142.

Oh, D., Buck, E. A., & Todorov, A. (2019). Revealing hidden gender biases in competence impressions of faces. *Psychological Science, 30*, 65–79.

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*, 323–342.

Oh, D., Shafir, E., & Todorov, A. (2020). Economic status cues from clothes affect perceived competence from faces. *Nature Human Behavior, 4*, 287–293.

Olivola, C. Y., & Todorov, A. (2010a). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior, 34*, 83–110.

Olivola, C. Y., & Todorov, A. (2010b). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology, 46*, 315–324.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*, 11087–11092.

Porter, S., England, L., Juodis, M., Ten Brinke, L. & Wilson, K. (2008). Is the face the window to the soul?: Investigation of the accuracy of intuitive judgments of the trustworthiness of human faces. *Canadian Journal of Behavioural Science, 40*, 171–177.

Poutvaara, P., Jordahl, H., & Berggren, N. (2009). Faces of politicians: Babyfacedness predicts inferred competence but not electoral success. *Journal of Experimental Social Psychology*, *45*, 1132–1135.

Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PLoS ONE*, *7*, e34293.

Rule, N. O., Ambady, N., & Adams, R. B. (2009). Personality in perspective: Judgmental consistency across orientations of the face. *Perception*, 38(11), 1688–1699.

Said, C. P., & Todorov, A. (2011). A statistical model of facial attractiveness. *Psychological Science*, *22*, 1183–1190.

Secord, P. F. (1958). Facial features and inference processes in interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior* (pp. 300–315). Stanford University Press.

Shen, X., Mann, T. C., & Ferguson, M. J. (2020). Beware a dishonest face? How we update our implicit impressions of untrustworthy faces. *Journal of Experimental Social Psychology*, 86, 103888.

Sussman, A. B., Petkova, K., & Todorov, A. (2013). Competence ratings in US predict presidential election outcomes in Bulgaria. *Journal of Experimental Social Psychology*, *49*, 771–775.

Todorov, A. (2017). *Face value: The irresistible influence of first impressions*. Princeton University Press.

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion*, *13*, 724–738.

Todorov, A., Dotsch, R., Wigboldus, D. H. J., & Said, C. P. (2011). Data-driven methods for modeling social perception: Modeling social perception. *Social and Personality Psychology Compass*, *5*, 775–791.

Todorov, A., Loehr, V., & Oosterhof, N. N. (2010). The obligatory nature of holistic processing of faces in social judgments. *Perception*, 39, 514–532.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, 308, 1623–1626.

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, 519–545.

Todorov, A., & Oosterhof, N. (2011). Modeling social perception of faces. *IEEE Signal Processing Magazine*, *28*, 117–122.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, *27*, 813–833.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*, 455–460.

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, 83, 1051–1065.

Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology*, 39, 549–562.

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, 87, 482–493.

Todorov, A., & Oh, D. (2021). The structure and perceptual basis of social judgments from faces. In *Advances in experimental social psychology* (Vol. 63, pp. 189–245). Academic Press.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 28, pp. 211–279). San Diego, CA: Academic Press.

Verosky, S. C., Porter, J. M., Martinez, J. E., & Todorov, A. (2018). Robust effects of affective person learning on evaluation of faces. *Journal of Personality and Social Psychology*, *114*, 516–528.

Verosky, S. C., & Todorov, A. (2010). Generalization of affective learning about faces to perceptually similar faces. *Psychological Science*, *21*, 779–785.

Verosky, S. C., & Todorov, A. (2013). When physical similarity matters: Mechanisms underlying affective learning generalization to the evaluation of novel faces. *Journal of Experimental Social Psychology*, *49*, 661–669.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*, 592–598.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*, 237–252. Also see correction in *Journal of Personality and Social Psychology* (1986), *50*, 355.

Zebrowitz McArthur, L., & Apatow, K. (1984). Impressions of baby-faced adults. *Social Cognition*, *2*, 315–342.