



Reputation in social perception: Validation of trustworthy and untrustworthy behavioural descriptions in European Portuguese

Marta Granadeiro* / Isabel Pavão Martins** / Leonel Garcia-Marques***

* Universidade de Lisboa, Faculdade de Medicina, Centro de Estudo Egas Moniz, Laboratório de Estudos de Linguagem, Lisboa, Portugal / Instituto Gulbenkian de Medicina Molecular (GIMM), Lisboa, Portugal; ** Universidade de Lisboa, Faculdade de Medicina, Centro de Estudo Egas Moniz, Laboratório de Estudos de Linguagem, Lisboa, Portugal / ULS Santa Maria, Departamento de Neurociências, Consulta de Cefaleias, Lisboa, Portugal; *** Universidade de Lisboa, Faculdade de Psicologia, Centro de Investigação em Ciência Psicológica (CICPSI), Lisboa, Portugal

Abstract: Reputation is an important cue for person perception and adequate behaviour. Its influence on behaviour might override first impressions from faces and constitutes an important information for integrating person perception and creating predictions of other's behaviour. The purpose of the present study was to create and validate a set of sentences describing "trustworthy" and "untrustworthy" previous behaviour (reputation). We generated a set of 97 sentences that were divided into three groups: trustworthy, untrustworthy and neutral. Trustworthy sentences elicited higher perceived trustworthiness compared to both neutral and untrustworthy sentences. Untrustworthy sentences elicited lower perceived trustworthiness compared to neutral sentences. This material seems to be effective in manipulating perceived trustworthiness and could be used as stimuli in the study of Spontaneous Trait Inferences.

Keywords: Social cognition, Social trait inference, Spontaneous trait inferences, Trustworthiness.

Introduction

"(...) The individual untrained in the pitfalls of personality appraisal, however, seldom hesitates or lacks confidence in his judgments of people when asked to make evaluations on the basis of the scantiest of information." (Secord et al., 1960).

The study of impression formation in person perception has begun more than seven decades ago. Impression formation is generally considered to be a fast and automatised (Willis & Todorov, 2006) process which has an unclear relationship with the real presence of the perceived traits (Todorov, 2017; Todorov et al., 2008; Zebrowitz, 2017; Zebrowitz & Montepare, 2008). This interesting ability is especially important because it has shown to predict important real outcomes such as political elections (Ballew & Todorov, 2007; Tigue et al., 2012), leader selection (Klofstad et al., 2012), economic decisions (Montano et al., 2017; O'Callaghan et al., 2016; Rezlescu et al., 2012), vocal pitch variation during daily conversations (Michalsky & Schoormann, 2017), job selection and legal decisions (Todorov, 2017; Zebrowitz & McDonald, 1991). Judgements of specific social traits

Correspondence concerning this article should be addressed to: Marta Granadeiro, Laboratório de Estudos de Linguagem, Centro de Estudo Egas Moniz, Faculdade de Medicina, Universidade de Lisboa, Avenida Professor Egas Moniz, 1649-028 Lisboa, Portugal. E-mail: marta.granadeiro@edu.ulisboa.pt

are strongly correlated, which makes it difficult to establish that the perception of a specific trait is leading to a certain behaviour (Todorov et al., 2008). Thus, a two-dimensional trait space has been proposed to reduce the amount of judgements from different social traits to two dimensions (McAleer et al., 2014; Todorov et al., 2008). This dimensional trait space is thought to represent the structure of social trait inference. The first dimension is associated with valence/trustworthiness and the second dimension is associated with power/competence/dominance (Todorov et al., 2008; Todorov & Oh, 2021) (also see Oliveira et al., 2019) for a discussion on the differences between perceived dominance and competence). Although this dimensional trait space summarises the relationship between different traits very elegantly, it has been shown that the predictive power of each perceived trait on behaviour also depends on its relative importance for the perceiver (Hall et al., 2009; Todorov, 2017). Meaning that the predictive power of a judgment on behaviour is expected to be greater when the trait is considered important for the specific context. For this reason, the study of the perception of individual traits is still important as a predictor of specific behaviours.

Perceived trustworthiness has been extensively studied in person perception as it is thought to be crucial for cooperation (although see Cook et al., 2005). Trust is often defined as performing an initial sacrifice that depending on other's response might be detrimental to the one that is self sacrificing (Alós-Ferrer & Farolfi, 2019). It exists when one party to the relation believes the other party has incentive to act in his or her interest or to take his or her interests into account (Cook et al., 2005). Although trust and trustworthiness are overlapping concepts, they seem to be determined by different factors. While trustworthy people tend to be more trusting, people more trusting are not necessarily trustworthy (Alós-Ferrer & Farolfi, 2019).

Particularly interesting is the effect of perceived trustworthiness on economic decision-making. One of the most used tasks for the study of trust-related decision-making are investment games. On such economic games, one player (A) starts with an initial endow and decides if they want to invest an amount of that initial endow on another player (B). The amount invested to the second player is multiplied by a factor (k) and sent. Player B decides how much of that money they want to keep and how much they send back (reciprocate) to player A (Berg et al., 1995). Game theory predicts that the best choice for player B is to keep all the money, therefore the best choice for player A is to send zero money in the first place. Despite this, the majority of people, playing as player A and B, send a part of the initial endow and reciprocate a part of the money received, respectively. Trust and trustworthiness are thought to play an important role in this behaviour. Specifically, participants invest more money when they perceive the other player as more trustworthy (Rezlescu et al., 2012; Van 't Wout & Sanfey, 2008).

The perception of trustworthiness has been studied using different types of stimuli. Faces are one of the most widely used stimuli in person perception. Literature suggests that faces that resemble that of babies (e.g., large eyes, rounder faces, feature placement more concentrated in the lower part of the face) and faces that resemble familiar or close relatives are judged as more trustworthy (Zebrowitz, 2017; Zebrowitz et al., 2003). (Dotsch & Todorov, 2012), through a reverse correlation method, found that the facial features more diagnostic of facial judgements were the eyes/eyebrows, mouth and hair region. Trustworthiness judgements inferred from voices have also shown to be very fast (one word seems enough) and to be relatively independent of language (Baus et al., 2019; McAleer et al., 2014). Also, sex differences have been described, with women investing more money in higher-pitched male voices (Montano et al., 2017), which in turn are associated with more perceived trustworthiness. A similar two-dimensional trait space has been described for voices, with warmth/trustworthiness and dominance as the two main dimensions (McAleer et al., 2014). Studies that use the integration of faces and voices have also suggested an interaction between the two types of stimuli. For attractiveness judgements, faces seem to have a prevalent importance but for dominance, voices seem to be more important

(Rezlescu et al., 2015). Despite the majority of the research on the perception of trustworthiness has been through the study of facial or acoustic cues, information from past behaviour (or reputation) has been shown important effects on social behaviour that may override that of facial cues (Rezlescu et al., 2012).

Reputation, i.e., information about a person's past behaviour, is of particular interest when studying the effect of social perception on behaviour. It is thought to serve as an important cue for cooperation and norm compliance (Origgi et al., 2018). It can be seen as a "social credit", that the individual possesses, that exerts pressure on others to behave accordingly with it. For instance, if person A is kind to person B, because person B speaks about it with person C, person C will more probably treat person A kindly in the future. Person A's reputation influenced how they were treated (Origgi et al., 2018). Experimentally, reputation seems to almost override the effect of first impressions from faces in trust-related decision-making (Rezlescu et al., 2012). When participants were exposed only to a person's face, they invested more money when they judged the person as more trustworthy. Despite this, when participants were given information on the reputation of the person, participants invested more money in good reputation (vs. bad reputation), irrespective of facial trustworthiness judgements.

Theories of indirect reciprocity suggest that reputation plays a crucial role in human social behaviour (Buckholtz & Marois, 2012; Origgi et al., 2018). The aim of the present study was to create and validate a set of sentences describing trustworthy/untrustworthy behaviour (reputation), that could be used to manipulate perceived trustworthiness. An additional set of neutral-content sentences (sentences not diagnostic of trustworthiness, such as "*Olhou pela janela e viu que estava a chover / Looked out the window and saw that it was raining*") was also validated to be used as a control for reputation. To our knowledge, a set of descriptions implying, specifically, trustworthy/untrustworthy behaviour for the European Portuguese language still does not exist.

Validating material for the study of reputation in person perception could help shed light on (1) the mechanisms of integration of previous facial and vocal social trait judgements with novel congruent or incongruent reputation information, (2) the relative importance of reputation on social decision-making and (3) the understanding of behaviour directed at improving our reputation and its use for building social structure complexity in humans.

Methods

Participants

One hundred and twenty six undergraduate students were recruited to participate in this study. All provided informed consent, in accordance with the Declaration of Helsinki, which was mandatory to proceed to the experiment. Demographic data for the participants is shown in Table 1. Participants received course credits in exchange for their participation. Study was approved by the Ethics Committee of the Faculty of Medicine of the University of Lisbon.

Inclusion criteria was (a) having more than 18 years and (b) having normal or corrected vision. As this study intended to validate sentences in the European Portuguese language, participants' first language had to be European Portuguese. Additionally, as there are known cultural differences in social trait inference (Todorov & Oh, 2021), participants had to have Portuguese nationality as well. Both other nationality or other first language were exclusion criteria. Three participants were excluded because they did not finish the task. Thirteen participants fulfilled exclusion criteria and were not included in further analysis. Data acquired from the remaining one hundred and eleven participants was included in the analysis.

Table 1

Participant's demographic information

Demographics		Participants (<i>N</i> = 111)	Percentage
Age (years)	Mean (<i>SD</i>)	20.42 (6.09)	
	Range	18-53	
Education (<i>n</i>)	High school	88	79.2%
	Bachelor's degree	21	18.9%
	Master's degree	2	1.8%
Sex (<i>n</i>)	Feminine	96	86.5%
	Masculine	14	12.6%
	Rather not say	1	0.9%

Procedure

Behavioural descriptions of trustworthiness. Four participants were recruited, by word of mouth, as initial judges [mean age 26.3 (\pm 1.7 years), two male, mean education was 17 years and all held a Master's in Psychology]. All of them were naive to the goal of the study. They were asked to think about what they thought represented trustworthy or untrustworthy behaviour and to generate short sentences describing general examples of it. Trustworthy behaviour was defined as behaviour that they thought meant that the person could be trusted. Three judges generated 3 sentences each and one judge generated two sentences. Sentences were sent by the judges via e-mail. Each of the short sentences was used as template to generate further behaviour-describing sentences (Table 2).

Table 2

Template sentences created by the independent judges

Sentences generated		Trustworthiness
Judge I	Procura ajuda quando necessita	Trustworthy
	Preocupado com o bem-estar de quem o rodeia	Trustworthy
	Honesto quanto ao comportamento, seja favorável ou não	Trustworthy
Judge II	Cumpre sempre a sua palavra/promessa	Trustworthy
	É franco/frontal	Trustworthy
	Não partilha os segredos que lhe foram confiados com outras pessoas	Trustworthy
Judge III	Cumpre sempre as suas promessas	Trustworthy
	É transparente nas suas ações	Trustworthy
	Demonstra respeito e empatia por todos	Trustworthy
Judge IV	Quebra frequentemente os seus Compromissos/promessas	Untrustworthy
	As suas atitudes são inconsistentes	Untrustworthy

Sixty one sentences describing trustworthy and untrustworthy behaviour were generated as variations of the content described in the short sentences (e.g., short sentence “Cumpre sempre a sua palavra” – variation “Prometeu que levaria a mãe a Fátima e cumpriu” or “Prometeu que cuidaria da filha e não cumpriu”. Twelve neutral-content (i.e., describing a behaviour that was not diagnostic of trustworthiness) sentences were included to assess bias in using the Likert scale and to serve as control in tasks manipulating the trustworthiness/untrustworthiness of sentences. Also, to better account for sentence equivalency (number of words, number of characters and sentence length) we generated 24 more sentences that were trustworthy, untrustworthy or neutral equivalents of the original sentences. These sentences have variations of word order or are negative versions of sentences that change the meaning of the sentence, therefore, changing trustworthiness impression (Table 3). Total number of sentences was ninety seven. Each sentence was numbered

and all the numbers were randomised using random.org software (Haahr, 1998) and it resulted in a single order of sentences. The sentences were presented in that order to all participants. All generated sentences were included in the validation task. The four initial judges did not participate in the validation of sentences.

Table 3

Sentences equivalents (maintaining similar syntax, word count and lenght)

Original	Trustworthiness	Equivalent	Trustworthiness
Na primeira semana do mês já tem que pedir dinheiro aos amigos	Untrustworthy	Na primeira semana do mês já consegue pagar o que deve aos amigos	Trusworthy
Fingiu que andava de muletas para poder passar à frente na fila	Untrustworthy	Andava de muletas e perguntou se podia para passar à frente na fila	Trusworthy
Não demonstrou afecto num momento em que um amigo tanto precisava	Untrustworthy	Demonstrou o seu afeto num momento em que um amigo precisava	Trusworthy
Ameaçou bater numa pessoa com quem estava a discutir	Untrustworthy	Nunca ameaçou ninguém com quem estava a discutir	Trusworthy
Incitou os colegas a fazer greve e foi o primeiro a desistir quando chegou o chefe	Untrustworthy	Incentivou os colegas a fazerem greve e foi o primeiro a defendê-los quando chegou o chefe	Trusworthy
Falou injustamente de um colega ao seu chefe, para receber mais comissões	Untrustworthy	Falou positivamente de um colega ao seu chefe e ele recebeu mais comissões	Trusworthy
Pela terceira vez, prometeu que ajudaria a tratar da filha e não cumpriu	Untrustworthy	Prometeu que ajudaria a tratar da filha e pela terceira vez cumpriu	Trusworthy
Passou à frente de três pessoas da fila sem pedir autorização a nenhuma	Untrustworthy	Deixou passar à frente três pessoas da fila que lhe pediram autorização	Trusworthy
Nem se lembrou dos colegas que o ajudaram no trabalho naquele dia	Untrustworthy	Naquele dia, lembrou-se dos seus colegas que o ajudaram no trabalho	Trusworthy
Deu troco a menos, como quem se tinha enganado, para ver se o cliente não notava	Untrustworthy	Notou que tinha dado troco a menos e disse ao cliente que se tinha enganado	Trusworthy
Encontra-se com amantes, mesmo tendo uma relação	Untrustworthy	Encontra-se com amigos com quem tem uma relação	Neutral
Nas cheias da Madeira, ajudou na evacuação de 23 pessoas	Trusworthy	Nas cheias da Madeira, assistiu à evacuação de 23 pessoas	Neutral
Passou uma semana inteira em casa do primo que tinha tido um acidente	Trusworthy	Passou uma semana inteira em casa do primo que tinha comprado um computador	Neutral
Contou aventuras que um amigo teve numa viagem a África como tendo sido suas	Untrustworthy	Durante o jantar, contou as aventuras que um amigo teve em África	Neutral
Consolou um amigo que perdeu um familiar	Trusworthy	Ignorou um amigo que perdeu um familiar	Untrustworthy
Durante as férias, alimentou todos os dias o gato de um colega	Trusworthy	Durante as férias, fingiu que alimentava o gato de um colega	Untrustworthy
Cedeu a sua vacina a uma criança que estava muito doente	Trusworthy	Tirou a vacina a uma criança que estava muito doente	Untrustworthy
Guardou um segredo	Trusworthy	Guardou um suborno	Untrustworthy
Encontrou uma carteira no chão e entregou à polícia	Trusworthy	Encontrou uma carteira no chão e fugiu da polícia	Untrustworthy
A caminho de casa, ofereceu o seu jantar a um sem-abrigo	Trusworthy	A caminho de casa, pontapeou o jantar de um sem-abrigo	Untrustworthy
Nunca traiu ninguém	Trusworthy	Nunca confiou em ninguém	Untrustworthy
Ajudou um homem a encontrar o filho no centro comercial	Trusworthy	Empurrou um homem que estava com o filho no centro comercial	Untrustworthy
Viu um acidente de carro e correu para ajudar os feridos	Trusworthy	Viu um acidente de carro e fugiu sem ajudar os feridos	Untrustworthy
Aceitou substituir um colega do trabalho que queria visitar a avó no hospital	Trusworthy	Negou substituir um colega do trabalho que queria visitar a avó no hospital	Untrustworthy

Validation task. The task was conducted online, using Qualtrics Survey Software. The instructions stated that behaviour of several people would be presented, they should read each sentence and rate, in a 11-point Likert scale, ranging from 1 [not at all trustworthy] to 11 [very trustworthy], how trustworthy they thought that person was. Participants were instructed to answer according to their first impression, although not having to answer within a specific time limit and were told there were no correct or incorrect answers. Participant’s age, sex, education level, nationality and first language was also collected. Total task duration was 10 minutes.

Statistical analysis

To assess inter-rater reliability in the inference of trustworthiness, Intraclass Correlation Coefficient (ICC) estimates and its 95% confidence intervals were calculated using R (R Core Team, 2024) package (“irr”) (Gamer et al., 2005), function “icc” based on a mean-rating (k = 11), absolute agreement, two-way random-effect model (McGraw & Wong, 1996). Conventionally, ICC values bellow 0.5 are considered to represent poor reliability, values between 0.5 and 0.75 are considered to represent moderate reliability, values between 0.75 and 0.9 are considered to represent good reliability and values above 0.9 are considered to represent excellent reliability (Koo & Li, 2016).

In order to compare trustworthiness ratings between the three groups of sentences (untrustworthy, neutral and trustworthy), we performed a linear mixed-effect model analysis of perceived trustworthiness as a function of sentence group. We used the R package (“lme4”), function “lmer” (Bates et al., 2015), with a random slope and intercept for rater and a random intercept for sentence. Sentence group was set as a fixed effect. Estimation method used was Maximum Likelihood (ML). Three outliers were identified using Cook’s D and excluded from the analysis. The model was refitted without those observations. *P*-values were obtain with the likelihood ratio tests of the full model against the model without the effect of sentence group.

Power analysis was performed using package “simr” (Green & MacLeod, 2016), function “powerCurve” (200 simulations), in order to determine the minimum amount of raters needed to reach, at least 80% power, which is usually considered adequate.

Results

Sentences were divided in three groups (trustworthy, untrustworthy and neutral) and perceived trustworthiness was compared between groups (Figure 1). Sentence group affected perceived trustworthiness [$\chi^2(2) = 198.83, p < 0.001$] (Table 4), lowering it by about 3.03 points \pm 0.28 (standard errors) from neutral to untrustworthy sentence group and increasing it by about 2.26 points \pm 0.27 (standard errors) from neutral to trustworthy sentence group (Table 5).

Table 4
ANOVA table for the effect of sentence group

Model	AIC	BIC	logLik	Chisq	df	<i>p</i> -value
Null model	38179	38245	-19081			
Full model	37984	38065	-18981	198.83	2	< 0.001

Table 5

Linear mixed-effects model for mean perceived trustworthiness for each group of sentences

Predictors	Estimates	SE	df	<i>t</i> value	<i>p</i>
Intercept	6.04	0.23	119.85	25.80	< 0.001
Untrustworthy sentence	-3.03	0.28	126.18	-10.64	< 0.001
Trustworthy sentence	2.26	0.27	118.55	8.27	< 0.001
Random effects	Effect	Variance	<i>SD</i>	Correlation	
Sentence	Intercept	0.76	0.87		
Rater	Intercept	0.67	0.82		
	Trustworthy	0.86	0.93	-0.20	
	Untrustworthy	1.23	1.11	-0.72	-0.37
Residual		1.81	1.35		

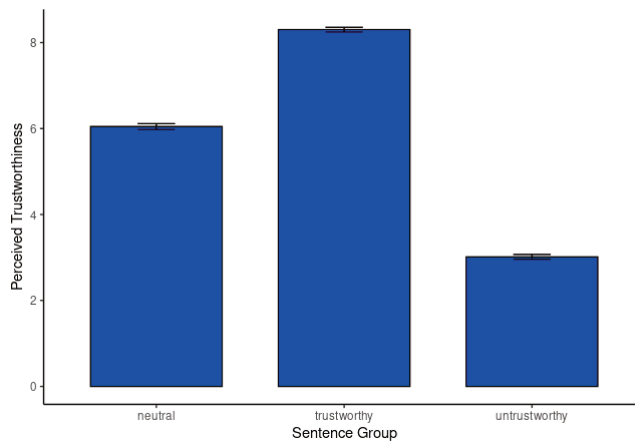


Figure 1. Estimated marginal means for each sentence group

Note. * $p < .001$.

Inter-rater reliability was excellent with ICC = 0.996 with a 95% CI = [0.995 - 0.997].

Power analysis indicated that maintaining number of sentences (97 sentences) and the effect size of the model, three raters would be enough to reach a 97% probability of finding an effect. This suggested that sample size was adequate to detect the effect.

The mean and standard deviation of the perceived trustworthiness of each sentence was also provided for the sake of transparency and in order to help in the selection of subsets of these sentences for different tasks (Appendix 1).

Discussion

In this study we aimed to validate a set of sentences that could manipulate perceived trustworthiness. We created and validated a set of sentences implying trustworthy and untrustworthy behaviour. Sentences implying trustworthy behaviour elicited higher judgements of perceived trustworthiness compared to sentences implying both neutral and untrustworthy behaviour. Similarly, neutral-behaviour sentences elicited higher judgements of perceived

trustworthiness compared to untrustworthy sentences. Trustworthiness judgements were highly similar between raters, with an ICC reporting excellent inter-rater reliability.

As expected, sentences implying trustworthy behaviour were rated as higher in trustworthiness people and sentences implying untrustworthy behaviour were rated as lower in untrustworthy. However, it is important to note that not all descriptions of trustworthy behaviour were rated as high in trustworthiness. This might be due to some descriptions implying more trustworthiness than others, meaning that some descriptions might be considered more diagnostic of trustworthiness than others. One example would be “Não contou a nenhum dos seus colegas que o seu pai esteve na prisão no passado / *Didn't tell any of their colleagues that his father was in prison in the past*”. This description was supposed to imply that the person is able to keep a secret, hence is trustworthy. Nevertheless, it may seem that the description refers more to hiding a secret than to keeping it, so its trustworthiness ratings were lower than expected. This also applies to some descriptions of untrustworthy behaviour. Both these behavioural descriptions had ratings of perceived trustworthiness that tended to get closer to that of neutral descriptions of behaviour (i.e., behaviour that is not related to a person's trustworthiness). Despite this, none of the descriptions of trustworthy behaviour were rated as untrustworthy and none of the descriptions of untrustworthy behaviour were rated as trustworthy. A set of descriptions with a combination of the highest/lowest mean perceived trustworthiness and lowest standard deviation should maximise the effect of sentence group on perceived trustworthiness.

Another important issue concerns assuring that the behavioural descriptions are text-based and not word-based. Some behavioural descriptions might just have specific words that imply specific traits (word-based inference), so the inference being made may not be the result of understanding the behaviour described (Orghian et al., 2018). In this study we wanted the content of the descriptions, as a whole, to drive trustworthiness inferences (text-based inferences), which was not objectively controlled. Nonetheless, this study only includes the inference of one trait (trustworthiness) and in many descriptions the only element that makes it trustworthy or untrustworthy is the presence of the negative word *no* (“*não*” in Portuguese). For instance, the sentence “*Prometeu que levaria a mãe a Fátima e cumpriu*” and the sentence “*Pela terceira vez, prometeu que ajudaria a tratar da filha e não cumpriu*” have very similar words but they have very different (almost opposite) mean trustworthiness ratings. Differentiating between the two sentences would demand a text-based inference, as changing the order of words or the presence of the negative word *no*, seemed to completely change perceived trustworthiness.

Reputation is a ubiquitous aspect of social interaction, that some argue is uniquely human (Emler, 1990; Izuma, 2012). It has an important role in social behaviour and is intimately related to cooperation and trustworthiness perception. ‘Good’ reputation elicits higher monetary investments which signals higher trustworthiness (King-Casas et al., 2005; Rezlescu et al., 2012). Additionally, the effects of reputation can override learning from trial and error, even when expectations driven by reputation were frequently violated (Delgado et al., 2005). The lack of reliance on feedback processing due to the interference of reputation information translated to an alteration of the neural pattern associated with feedback learning. Similarly, in economic-decision making, reputation information almost completely overrides the effect of perceived facial trustworthiness in the amount of money invested in others (Rezlescu et al., 2012). The study of reputation in person perception and decision-making is of high importance and provides a more ecological perspective of social behaviour as people often interpret other's reputation and make efforts for their own reputation (Emler, 1990).

In conclusion, the set of behavioural descriptions (reputation) seems to be effective in manipulating perceived trustworthiness. To our knowledge, a set of validated reputation sentences for the manipulation of social trustworthiness does not exist in European Portuguese. We believe person perception and decision-making are strongly influenced by reputation information,

sometimes at the expense of other relevant social information (i.e., faces or directly observed behaviour) and, therefore, material conveying this information is of importance.

Funding

This study was supported by Fundação para a Ciência e a Tecnologia by the research grant 2021.04582.BD.

Informed consent

All participants in this study were informed of the purpose of the study and how data will be used. They were assured that their identities would remain anonymous across the study.

Declaration of conflicting of interests

The author(s) declares no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. None of the authors has financial, personal, or professional affiliations that could influence the objectivity or integrity of the research process or the reporting of the findings.

Ethical approval

The study was approved by the Ethics Committee of the Faculty of Medicine of the University of Lisbon.

Author contributions

Conceptualization: LGM; Methodology: LGM; Formal analysis: MG, LGM; Funding acquisition: MG; Data curation: MG; Software: MG; Investigation: MG; Resources: MG; Visualization: MG; Validation: LGM; Supervision: IPM, LGM; Writing – Original draft: MG; Writing – Review & Editing: IPM, LGM.

All authors have read and agreed to the published version of the manuscript.

References

- Alós-Ferrer, C., & Farolfi, F. (2019). Trust games and beyond. *Frontiers in Neuroscience*, 13, e887. <https://doi.org/10.3389/fnins.2019.00887>
- Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences*, 104(46), 17948-17953. <https://doi.org/10.1073/pnas.0705435104>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>
- Baus, C., McAleer, P., Marcoux, K., Belin, P., & Costa, A. (2019). Forming social impressions from voices in native and foreign languages. *Scientific Reports*, 9(1), e414. <https://doi.org/10.1038/s41598-018-36518-6>

- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1), 122-142. <https://doi.org/10.1006/game.1995.1027>
- Buckholtz, J. W., & Marois, R. (2012). The roots of modern justice: Cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience*, 15(5), 655-661. <https://doi.org/10.1038/nn.3087>
- Cook, K. S., Hardin, R., & Levi, M. (2005). *Cooperation without trust?*. Russell Sage Foundation.
- Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, 8(11), 1611-1618. <https://doi.org/10.1038/nn1575>
- Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, 3(5), 562-571. <https://doi.org/10.1177/1948550611430272>
- Emler, N. (1990). A social psychology of reputation. *European Review of Social Psychology*, 1(1), 171-193. <https://doi.org/10.1080/14792779108401861>
- Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2005). *irr: Various coefficients of interrater reliability and Agreement* (Version 0.84.1) [R package]. CRAN. <https://doi.org/10.32614/CRAN.package.irr>
- Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493-498. <https://doi.org/10.1111/2041-210X.12504>
- Haahr, M. (1998). *RANDOM.ORG: True random number service*. <https://www.random.org>
- 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, C. M. Federico, & J. L. Sullivan (Eds.), *The Political Psychology of Democratic Citizenship* (1st ed., ch. 4, pp. 73-99). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195335453.003.0004>
- Izuma, K. (2012). The social neuroscience of reputation. *Neuroscience Research*, 72(4), 283-288. <https://doi.org/10.1016/j.neures.2012.01.003>
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science, New Series*, 308(5718), 78-83.
- Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B: Biological Sciences*, 279(1738), 2698-2704. <https://doi.org/10.1098/rspb.2012.0311>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- McAleer, P., Todorov, A., & Belin, P. (2014). How do you say 'hello'? Personality impressions from brief novel voices. *PLoS ONE*, 9(3), e90779. <https://doi.org/10.1371/journal.pone.0090779>
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30-46. <https://doi.org/10.1037/1082-989X.1.1.30>
- Michalsky, J., & Schoormann, H. (2017). Pitch convergence as an effect of perceived attractiveness and likability. *Interspeech 2017*, 2253-2256. <https://doi.org/10.21437/Interspeech.2017-1520>
- Montano, K. J., Tigue, C. C., Isenstein, S. G. E., Barclay, P., & Feinberg, D. R. (2017). Men's voice pitch influences women's trusting behavior. *Evolution and Human Behavior*, 38(3), 293-297. <https://doi.org/10.1016/j.evolhumbehav.2016.10.010>
- O'Callaghan, C., Bertoux, M., Irish, M., Shine, J. M., Wong, S., Spiliopoulos, L., Hodges, J. R., & Hornberger, M. (2016). Fair play: Social norm compliance failures in behavioural variant frontotemporal dementia. *Brain*, 139(1), 204-216. <https://doi.org/10.1093/brain/awv315>

- Oliveira, M., Garcia-Marques, T., Dotsch, R., & Garcia-Marques, L. (2019). Dominance and competence face to face: Dissociations obtained with a reverse correlation approach. *European Journal of Social Psychology*, 49(5), 888-902. <https://doi.org/10.1002/ejsp.2569>
- Orghian, D., Ramos, T., Reis, J., & Garcia-Marques, L. (2018). Acknowledging the role of word-based activation in spontaneous trait inferences. *Análise Psicológica*, 36(1), 115-131. <https://doi.org/10.14417/ap.1319>
- Origgi, G., Holmes, S., & Arikha, N. (2018). *Reputation: What it is and why it matters* (Eng. ed.). Princeton University Press.
- R Core Team. (2024). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- 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(3), e34293. <https://doi.org/10.1371/journal.pone.0034293>
- Rezlescu, C., Penton, T., Walsh, V., Tsujimura, H., Scott, S. K., & Banissy, M. J. (2015). Dominant voices and attractive faces: The contribution of visual and auditory information to integrated person impressions. *Journal of Nonverbal Behavior*, 39(4), 355-370. <https://doi.org/10.1007/s10919-015-0214-8>
- Secord, P. F., Stritch, T. F., & Johnson, L. (1960). The role of metaphorical generalization and congruency in the perception of facial characteristics. *The Journal of Social Psychology*, 52(2), 329-337. <https://doi.org/10.1080/00224545.1960.9922088>
- Tigue, C. C., Borak, D. J., O'Connor, J. J. M., Schandl, C., & Feinberg, D. R. (2012). Voice pitch influences voting behavior. *Evolution and Human Behavior*, 33(3), 210-216. <https://doi.org/10.1016/j.evolhumbehav.2011.09.004>
- Todorov, A. (2017). *Face value: The irresistible influence of first impressions*. Princeton University Press.
- Todorov, A., & Oh, D. (2021). The structure and perceptual basis of social judgments from faces. In B. Gawronski (Ed.), *Advances in experimental social psychology* (Vol. 63, pp. 189-245). Elsevier. <https://doi.org/10.1016/bs.aesp.2020.11.004>
- Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, 12(12), 455-460. <https://doi.org/10.1016/j.tics.2008.10.001>
- Van 'T Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, 108(3), 796-803. <https://doi.org/10.1016/j.cognition.2008.07.002>
- 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>
- Zebrowitz, L. A. (2017). First impressions from faces. *Current Directions in Psychological Science*, 26(3), 237-242. <https://doi.org/10.1177/0963721416683996>
- Zebrowitz, L. A., Fellous, J.-M., Mignault, A., & Andreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: Evidence from connectionist modeling. *Personality and Social Psychology Review*, 7(3), 194-215. https://doi.org/10.1207/S15327957PSPR0703_01
- Zebrowitz, L. A., & McDonald, S. M. (1991). The impact of litigants' baby-facedness and attractiveness on adjudications in small claims courts. *Law and Human Behavior*, 15(6), 603-623. <https://doi.org/10.1007/BF01065855>
- Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matter. *Social and Personality Psychology Compass*, 2(3), 1497-1517. <https://doi.org/10.1111/j.1751-9004.2008.00109.x>

Appendix 1

Mean and standard deviation of perceived trustworthiness for each sentence

Group	Sentences	Mean	SD
Trustworthy	Não deixou nenhum colega para trás quando foram atacados na guerra do Ultramar	9,68	1,62
Trustworthy	Guardou um segredo	9,42	1,64
Trustworthy	Cedeu a sua vacina a uma criança que estava muito doente	9,33	1,71
Trustworthy	Encontrou uma carteira no chão e entregou à polícia	9,32	1,44
Trustworthy	Nas cheias da Madeira, ajudou na evacuação de 23 pessoas	9,13	1,74
Trustworthy	Prometeu que levaria a mãe a Fátima e cumpriu	8,99	1,96
Trustworthy	Durante as férias, alimentou todos os dias o gato de um colega	8,91	1,61
Trustworthy	Incentivou os colegas a fazerem greve e foi o primeiro a defendê-los quando chegou o chefe	8,87	2,06
Trustworthy	Disse sempre a verdade quando foi testemunhar em tribunal	8,83	1,60
Trustworthy	Viu um acidente de carro e correu para ajudar os feridos	8,82	1,81
Trustworthy	Ofereceu-se para pagar a consulta de uma mãe doente e com dificuldades económicas	8,81	1,79
Trustworthy	Faz, todas as semanas, as compras para a sua vizinha de 90 anos	8,81	1,60
Trustworthy	Consolou um amigo que perdeu um familiar	8,78	1,82
Trustworthy	Aceitou substituir um colega do trabalho que queria visitar a avó no hospital	8,77	1,67
Trustworthy	Passou uma semana inteira em casa do primo que tinha tido um acidente	8,66	1,58
Trustworthy	Não se importou de dormir na sala para alojar bem as suas visitas	8,65	1,78
Trustworthy	Nunca traiu ninguém	8,62	2,03
Trustworthy	Convidou para o pódio, o seu adversário que perdeu por se ter lesionado	8,60	2,06
Trustworthy	Ajudou um homem a encontrar o filho no centro comercial	8,60	1,70
Trustworthy	Faz dádiva de sangue de 4 em 4 meses, desde há vários anos	8,60	1,89
Trustworthy	Demonstrou o seu afeto num momento em que um amigo precisava	8,59	1,53
Trustworthy	Ajudou o filho de uns vizinhos com os trabalhos de casa, enquanto os pais foram passear	8,56	1,67
Trustworthy	A caminho de casa, ofereceu o seu jantar a um sem-abrigo	8,56	1,76
Trustworthy	Notou que tinha dado troco a menos e disse ao cliente que se tinha enganado	8,55	1,97
Trustworthy	Todos os anos contribui com o que pode para o banco alimentar	8,51	1,85
Trustworthy	Viu a apresentação de um colega que estava muito nervoso e elogiou-o de forma positiva no final	8,33	1,85
Trustworthy	Tornou-se enfermeiro para salvar vidas	8,32	1,80
Trustworthy	Ficou no escritório mais 1 hora, mesmo sabendo que não ganharia mais por isso	8,26	1,80
Trustworthy	Falou positivamente de um colega ao seu chefe e ele recebeu mais comissões	8,25	1,90
Trustworthy	Mudou os seus planos para se ajustar aos dos seus colegas	8,22	1,92
Trustworthy	É presidente de uma empresa e nunca aceitou subornos	8,12	1,83
Trustworthy	Fez os arranjos na entrada do seu prédio, depois de ver coisas estragadas	8,08	1,69
Trustworthy	Não julgou alguém de quem os outros estavam a falar mal	8,00	2,03
Trustworthy	Prometeu que ajudaria a tratar da filha e pela terceira vez cumpriu	7,88	2,37
Trustworthy	Na primeira semana do mês já consegue pagar o que deve aos amigos	7,78	1,95
Trustworthy	Deixou passar à frente três pessoas da fila que lhe pediram autorização	7,63	1,81
Trustworthy	Não se riu quando o amigo lhe contou o seu maior segredo	7,51	1,86
Trustworthy	Nunca ameaçou ninguém com quem estava a discutir	7,49	1,85
Trustworthy	Decidiu esquecer que a colega a prejudicou no exame e dar-lhe uma nova oportunidade	7,28	1,79
Trustworthy	Naquele dia, lembrou-se dos seus colegas que o ajudaram no trabalho	7,18	1,51
Trustworthy	Castigou os dois filhos da mesma forma quando estragaram a torneira da casa de banho	7,07	2,13
Trustworthy	Transportou os copos novos devagar até casa sem partir nenhum	7,02	1,62
Trustworthy	Andava de muletas e perguntou se podia para passar à frente na fila	6,14	1,98
Trustworthy	Não contou a nenhum dos seus colegas que o seu pai esteve na prisão no passado	5,63	1,29
neutral	Encontra-se com amigos com quem tem uma relação	6,68	1,88
neutral	Durante o jantar, contou as aventuras que um amigo teve em África	6,61	1,58
neutral	Comprou uma mesa para colocar na sala	6,40	1,55
neutral	Utiliza o metro para se deslocar para o trabalho	6,26	1,41
neutral	Pratica ginástica todas as terças e quintas	6,26	1,31
neutral	Quando foi a um restaurante, perguntou se tinham bitoque com arroz	6,17	1,49
neutral	Perto de sua casa, avistou um conjunto de patos a mergulharem num lago	6,02	1,24
neutral	Gosta de comer gelado de morango e baunilha, quando vai ao café da sua rua	5,99	1,27
neutral	No caminho para o centro comercial, passou por outra loja e comprou café	5,99	1,27
neutral	Olhou pela janela e viu que estava a chover	5,95	1,29
neutral	Encontrou o seu vizinho quando foi buscar o correio	5,95	1,20
neutral	Aproximou-se da multidão, com curiosidade	5,91	0,94
neutral	A sua cor preferida é o azul	5,89	1,42
neutral	Escreve com a mão direita	5,86	1,26
neutral	Nas cheias da Madeira, assistiu à evacuação de 23 pessoas	5,50	1,93
neutral	Passou uma semana inteira em casa do primo que tinha comprado um computador	5,24	1,05
Untrustworthy	Despediu-se do trabalho durante uma discussão com o chefe	5,57	1,24
Untrustworthy	Ganhou uma herança considerável e ainda diz querer ganhar o Euromilhões	5,23	1,75
Untrustworthy	Perguntou 3 vezes à recepcionista se faltava muito para ser atendida	5,07	1,49

Group	Sentences	Mean	SD
Untrustworthy	Perguntou algo que sabia de antemão para testar o seu colega	4,66	1,82
Untrustworthy	Mesmo tendo um salário reduzido, acreditava que ia conseguir comprar um carro desportivo a curto prazo	4,45	1,64
Untrustworthy	Nunca confiou em ninguém	4,42	1,86
Untrustworthy	Fez um comentário sobre alguém que se encontrava por perto e podia ouvir	4,08	1,66
Untrustworthy	Arranjou o exame dos outros anos mas não contou a ninguém	3,96	2,09
Untrustworthy	Bateu com a porta da sala porque queria uma nota melhor no teste	3,88	1,72
Untrustworthy	Descreveu o roubo que pescou como se fosse um tubarão	3,77	1,51
Untrustworthy	Aceitou trabalhar para o rival do seu melhor cliente a troco de mais algum dinheiro por mês	3,50	1,98
Untrustworthy	Não demonstrou afecto num momento em que um amigo tanto precisava	3,39	1,64
Untrustworthy	Na primeira semana do mês já tem que pedir dinheiro aos amigos	3,36	1,55
Untrustworthy	Nem se lembrou dos colegas que o ajudaram no trabalho naquele dia	3,19	1,38
Untrustworthy	É arguido num caso de roubo com arma branca	3,19	1,52
Untrustworthy	Negou substituir um colega do trabalho que queria visitar a avó no hospital	3,17	1,57
Untrustworthy	Está sempre a corrigir os outros quanto aos seus hábitos alimentares mas depois só come fritos	3,14	1,57
Untrustworthy	Ajuda clandestinamente amigos empresários, a troco de uma porção do lucro	2,98	1,72
Untrustworthy	Ameaçou bater numa pessoa com quem estava a discutir	2,81	1,60
Untrustworthy	Passou à frente de três pessoas da fila sem pedir autorização a nenhuma	2,74	1,46
Untrustworthy	Guardou um suborno	2,69	1,33
Untrustworthy	Contou aventuras que um amigo teve numa viagem a África como tendo sido suas	2,56	1,35
Untrustworthy	Viu um acidente de carro e fugiu sem ajudar os feridos	2,46	1,66
Untrustworthy	Encontrou uma carteira no chão e fugiu da polícia	2,31	1,64
Untrustworthy	Ignorou um amigo que perdeu um familiar	2,31	1,41
Untrustworthy	Deu troco a menos, como quem se tinha enganado, para ver se o cliente não notava	2,24	1,25
Untrustworthy	Incitou os colegas a fazer greve e foi o primeiro a desistir quando chegou o chefe	2,21	1,62
Untrustworthy	Fingiu que andava de muletas para poder passar à frente na fila	2,17	1,41
Untrustworthy	Apontou o x-acto a uma pessoa com quem estava a discutir	2,11	1,56
Untrustworthy	Falou injustamente de um colega ao seu chefe, para receber mais comissões	1,95	1,39
Untrustworthy	Empurrou um homem que estava com o filho no centro comercial	1,91	1,23
Untrustworthy	Passou a imagem que queria ajudar a colega quando o que queria era ficar com o seu cargo	1,89	1,40
Untrustworthy	Tirou a vacina a uma criança que estava muito doente	1,89	1,34
Untrustworthy	Pela terceira vez, prometeu que ajudaria a tratar da filha e não cumpriu	1,72	1,46
Untrustworthy	A caminho de casa, pontapeou o jantar de um sem-abrigo	1,58	1,26
Untrustworthy	Durante as férias, fingiu que alimentava o gato de um colega	1,54	1,20
Untrustworthy	Encontra-se com amantes, mesmo tendo uma relação	1,45	1,23

Reputação na percepção social: Validação de descrições de comportamento confiável e não confiável em Português Europeu

Resumo: A reputação é uma pista importante para a percepção de pessoas e adequação do comportamento. A sua influência no comportamento sobrepõe-se frequentemente à das primeiras impressões com base na face e constitui uma informação importante para integrar a percepção de pessoas e criar expectativas do comportamento de outros. O objetivo deste estudo foi criar e validar um conjunto de frases que descrevem comportamento prévio (reputação) “confiável” e “não confiável”. Gerámos um conjunto de 97 frases que foram divididas em três grupos: confiáveis, não confiáveis e neutras. As frases confiáveis elicitaram maior confiabilidade percebida em comparação com frases não confiáveis e neutras. Frases não confiáveis elicitaram menor confiabilidade percebida em comparação com frases neutras. Este material parece eficaz e adequado para a manipulação da confiabilidade percebida e poderá ser utilizado como estímulo no estudo da Inferência Espontânea de Traços.

Palavras-chave: Cognição social, Inferência de traços sociais, Inferência espontânea de traços, Confiabilidade.