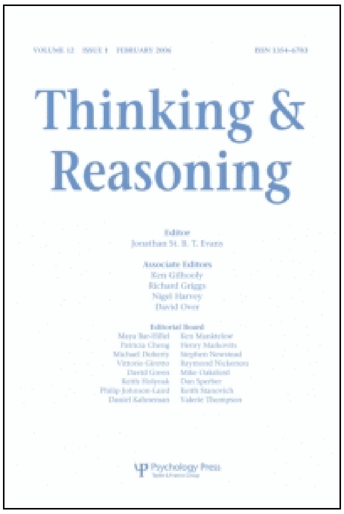

Thinking & Reasoning
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713685607

A judgement analysis of social perceptions of attitudes and ability

Brent Snook [a]; Malcolm J. Grant [a]; Cathryn M. Button [a]
[a] Memorial University, St. John's, NL, Canada

First Published on: 22 August 2009

To cite this Article Snook, Brent, Grant, Malcolm J. and Button, Cathryn M.(2009)'A judgement analysis of social perceptions of attitudes and ability',Thinking & Reasoning,15:4,319 — 336

To link to this Article: DOI: 10.1080/13546780903135722

URL: http://dx.doi.org/10.1080/13546780903135722

THINKING & REASONING, 2009, 15 (4), 319–336

Psychology Press
Taylor & Francis Group

A judgement analysis of social perceptions of attitudes and ability

Brent Snook, Malcolm J. Grant, and Cathryn M. Button
Memorial University, St. John's, NL, Canada

A judgement analysis of people's social inferences of attitudes and ability was conducted. University students were asked to infer the liberalness ($N = 60$; Study 1) or intelligence ($N = 40$; Study 2) of targets seen in pictures. Multiple regression analyses revealed that attractiveness was the most important cue for predicting inferences of liberalness, while an ethnic cue (i.e., being Asian) was the most important cue for judgements about intelligence. Results also showed that a single-cue model was less susceptible to overfitting, but significantly less accurate than a multiple-cue model in predicting participant's intelligence judgements. Although the multiple regression models suffered a degree of overfitting, cross validation showed that they continued to have significant predictive value when applied to new data. Furthermore, a "random partner" method (comparing each participant's own regression equation with that of another, randomly selected, participant) provided evidence of significant idiosyncratic variation in the way intelligence judgements were made.

Keywords: Abilities; Attitudes; Judgment analysis; Social inference; Social rationality.

Decades of research have shown that social perceptions of attitudes and abilities, particularly in first encounters, are influenced by physical cues such as age, gender, attractiveness, and ethnicity (Alley, 1988; Brewer, 1988; Brunswik, 1945; Bull & Rumsey, 1988; Dion, Berscheid, & Walster, 1972; Fiske & Neuberg, 1990; Grant, Button, Hannah, & Ross, 2002; Herman, Zanna, & Higgins, 1986; Rhodes & Zebrowitz, 2002). However, it is not clear if all people use the same cues and, if so, to what extent they do so

Correspondence should be address to Brent Snook, Department of Psychology, Memorial University of Newfoundland, St. John's, NL, Canada, A1B 3X9.
E-mail: bsnook@play.psych.mun.ca

Support of the research reported in this paper was provided to the first author by the Natural Sciences and Engineering Research Council of Canada.

http://www.psypress.com/tar DOI: 10.1080/13546780903135722

consistently over time and targets. In the present paper we applied multiple regression to the data of individual participants in order to explore these issues as they pertain to people's inferences about the attitude liberalness (Study 1) or intelligence (Study 2) of persons they saw in pictures. We chose to examine inferences about attitudes and intelligence because of the central role these domains play in a variety of social and competitive undertakings. Our liking for others, our readiness to trust them, to hire them, to challenge them, to admit them to our social group, and to try to influence them all depend to some degree on our assessments of their attitudes and abilities (e.g., Macrae & Bodenhausen, 2000).

Many social psychologists have assumed that people take advantage of the multiplicity of information available to them in forming their perceptions of others (see Eagly & Chaiken, 1993, for an overview of integration models). Specifically, researchers have often chosen to model social perception processes using models that combine various bits of information in an algebraic manner. Such integration models (e.g., Abelson & Levi, 1985; Anderson, 1982; Fishbein, 1980; Wyer, 1974) have a long history in social psychology and have frequently been shown to account for significant portions of the variability in people's judgements (e.g., Himmelfarb & Anderson, 1975). These models have also proven useful for understanding a wide range of judgement and decision-making processes (e.g., Dalgleish, 1988; Heald, 1991). However, the extent to which these models actually represent models of mind has been surrounded by much controversy (see Dhami & Harries, 2001; Doherty & Brehmer, 1997; Gigerenzer, Todd, & The ABC Research Group, 1999; Hammond, 1996; Hoffman, 1960). Notwithstanding this controversy, the wealth of research using multiple regression to understand human inferences suggests that the regression models will successfully capture our participants' inferences about the attitude liberalness and intelligence of others. In this research, we use multiple regression to learn about the relationships between physical cues and social inferences, but make no claim that the resulting models are isomorphic with the cognitive processes underlying people's judgements.

One of the most widely studied physical cues in social inferences has been attractiveness. Several reviews (e.g., Eagly, Ashmore, Makhijani, & Longo, 1991; Hatfield & Sprecher, 1986) have shown that attractiveness is a characteristic that affects many different types of evaluative social inferences (i.e., "What is beautiful is good"). However, people's willingness to make judgements about others based on their attractiveness may vary according to the type of inference they are asked to make. For instance, Eagly and her colleagues (1991) found that attractiveness accounted for more of the variability in judgements about social competence than in judgements about intelligence, potency, and adjustment. It appears that people are less willing to use physical cues, such as attractiveness, when making inferences about someone's abilities, but are more willing to do so when making inferences about attitudes.

Although age and gender also influence many social perceptions, the direction of the inference using these cues is usually more context specific than is the case with attractiveness. For example, women and young people are thought to be more liberal (Grant, Button, Ross, & Hannah, 1997; Grant, Ross, Button, Hannah, & Hoskins, 2001) but gender and age are not so clearly related to judgements of intelligence. How other clearly visible characteristics such as ethnicity, degree of smiling, and cultural signifiers (e.g., traditional dress, facial markings, etc.), are related to social judgements remains to be determined.

The extent to which there is agreement regarding the emphasis placed on different physical cues has not received much attention. Even if people agree that certain cues are relevant to a particular judgement, they may disagree about how those cues should be used or how many are required. In the present study we assessed cue use at the individual level in order to capture both idiosyncratic and consensual aspects of stereotypic judgements.

The usefulness of any model of judgement rests on its ability to perform well on data that were not used in the construction of the model (Gigerenzer et al., 1999). Models of judgement, especially those with many free parameters, tend to overfit the data. As a result, these models usually lose some of their predictive power when applied to novel data (Efron, 1982). Given this possibility, cross-validation of any judgement model is a necessary test of its validity. A variety of cross-validation techniques will be compared in the present paper.

In the current research we use multiple regression to examine the following issues: the cues that are related to social inferences and their weightings; the level of agreement across different participants with respect to the weightings given to the various cues; and the effect of judgement domain on the ability to predict people's social inferences.

STUDY 1

Method

Participants

Participants were 60 undergraduates (30 men and 30 women) at Memorial University.

Materials

Target pictures. Digitised pictures of adult men and women (showing head and shoulders) were obtained from a variety of websites. In total there were 143 women and 145 men. All pictures were digitally cropped to a width of 172 pixels and a height of 203 pixels and saved as 256-colour, bit-mapped images. When displayed on a participant's computer screen, the images were approximately 4.7 cm wide and 5.4 cm high.

All targets had been rated on apparent age (in years) and attractiveness (1 = *unattractive* to 10 = *attractive*), and categorised according to apparent ethnicity by participants in earlier studies (Grant et al., 2000). For the purpose of this study, ethnicity was broken into three binary cues: Asian or not, Black or not, and Caucasian or not.[1] These were categorisations on which the earlier participants had shown high levels of agreement. The presence or absence of cultural signifiers such as facial markings, headwear, etc., and the extent of smiling, were judged independently by two of the authors. On cultural signifiers, judges agreed on all but 3 of the 288 targets and disagreements were resolved following discussion. On smiling, judges rated each target on a 4-point scale: 1 (no smile), 2 (slight smile), 3 (moderate smile), and 4 (full smile). The correlation between judges' ratings was $r(286) = .86$, $p < .001$. Judges gave identical ratings to 182 of the targets (63%), ratings that differed by one scale point to 90 targets (31%) and ratings that differed by two scale points to the remaining 16 targets (6%). The rating assigned to the target was the average of the two judges' ratings. Table 1 contains summary statistics for the target variables and Table 2 contains the correlations among those variables.

Attitude statements. Two statements on each of five issues (discipline of children, homosexuality, feminism, immigration, and religion) were used. For each issue, one of the statements expressed a conservative position and one expressed a liberal position. Each statement was followed by a 7-point scale where 1 was labelled "Strongly disagree" and 7 was labelled "Strongly agree". Liberalness scores were calculated by reverse scoring responses on the five conservative statements and then adding all 10 responses together. Scores could thus range from 10 (very conservative) to 70 (very liberal). The set of attitude statements is shown in Table 3 along with statistics indicating the internal consistency of the statements when used to measure liberalness.

Procedure

Participant inferences. Participants were tested up to three at a time. Each person sat in a separate cubicle equipped with a personal computer running a *Visual Basic* program. All instructions and experimental materials were presented on the computer screen and participants responded by pointing and clicking the mouse. For each participant 20 pictures (10 men and 10 women) were randomly selected from the pool of 288 pictures.

[1] In this paper we use the term "Asian" to refer to persons who are apparently from East Asia, such as Chinese, Japanese, Koreans, etc.

TABLE 1
Cue properties for male and female targets of different ethnic origin used in Study 1 and Study 2

Ethnicity	*Age Mean (SD)*	*Attractiveness Mean (SD)*	*Smiling Mean (SD)*	*Cultural signifier (%)*
Asian				
Men ($n = 22$)	31.69 (5.64)	3.96 (0.87)	2.11 (1.27)	0.00
Female ($n = 29$)	29.92 (5.56)	5.44 (0.81)	2.43 (1.19)	3.45
Combined ($n = 51$)	30.68 (5.61)	4.80 (1.11)	2.29 (1.23)	1.96
Black				
Male ($n = 46$)	31.90 (6.05)	4.41 (0.70)	1.92 (1.15)	0.00
Female ($n = 64$)	31.04 (6.11)	5.00 (0.90)	2.65 (1.24)	0.00
Combined ($n = 110$)	31.40 (6.07)	4.76 (0.87)	2.35 (1.25)	0.00
Caucasian				
Male ($n = 35$)	40.61 (13.07)	3.96 (0.87)	2.19 (1.14)	0.00
Female ($n = 20$)	37.26 (11.22)	5.15 (1.34)	2.93 (1.17)	0.00
Combined ($n = 55$)	39.39 (12.43)	4.39 (1.20)	2.45 (1.20)	0.00
Other*				
Male ($n = 42$)	35.66 (7.58)	3.47 (0.72)	1.77 (1.08)	28.57
Female ($n = 30$)	43.78 (13.01)	3.34 (1.28)	2.22 (1.13)	83.33
Combined ($n = 72$)	39.04 (10.89)	3.42 (0.99)	1.96 (1.12)	51.39

*This category included East Indian ($n = 27$), Middle East ($n = 31$), Latin American ($n = 5$), and North American Native ($n = 9$).

TABLE 2
Correlations among the target cues used in Study 1 and Study 2

	Target cues						
	Gender	*Age*	*Attractive*	*Asian*	*Black*	*Caucasian*	*Smiling*
Age	−.04						
Attractiveness	.35**	−.61**					
Asian	.07	−.19**	.18**				
Black	.13*	−.27**	.27**	−.37**			
Caucasian	−.13*	.24**	.01	−.23**	−.38**		
Smiling	.24**	−.06	.33**	.01	.06	.08	
Cultural Signifier	.15*	.33**	−.41**	−.15**	−.31**	−.19**	−.08

The dichotomous variables were coded as follows: Gender (1 = male and 2 = female); Asian (0 = no and 1 = yes); Black (0 = no and 1 = yes); Caucasian (0 = no and 1 = yes); Cultural Signifier (1 = no and 2 = yes). In all cases, $n = 288$.
*$p < .05$ level (two-tailed). **$p < .01$ (two-tailed).

Pictures of these target persons were displayed, in a random order, one at a time on the participant's computer screen. While each picture remained on the screen, the 10 attitude statements were presented, in a random order, one

TABLE 3
Liberalism and conservatism attitude statements used in Study 1

Statement number	*Issue*	*Statement*	*Item-total correlation*	*Alpha if item deleted*
1	Discipline	A teacher should not be allowed to physically punish children. (L)	.153	.661
2	Homosexuality	School boards should not hire homosexual teachers. (C)	.479	.601
3	Immigration	It's time to close the door to refugees. (C)	.449	.598
4	Religion	Religious beliefs are important guiding principles in my life. (C)	.300	.632
5	Feminism	A woman's place is in the home. (C)	.201	.647
6	Homosexuality	Homosexuality should be accepted. (L)	.559	.598
7	Discipline	What young people need most of all is strict discipline by their parents. (C)	.387	.611
8	Immigration	I favour a more open immigration policy for Canada. (L)	.367	.615
9	Feminism	The feminist movement deserves strong support. (L)	.200	.648
10	Religion	Religion is mostly superstition. (L)	.229	.653

(L) indicates a liberal statement; (C) indicates a conservative statement. Statistics are based on participants' ratings of their own attitudes. Cronbach's alpha for the 10-item scale was .651.

at a time. The participant was asked to judge, using a 7-point scale that ranged from (1) strongly disagree to (7) strongly agree, how the person in the picture would respond to the attitude statement. When all 10 statements had been presented for a particular photograph, a new picture appeared and the procedure was repeated. The procedure ended when participants had made 10 attitude inferences for each of the 20 different target persons.[2] The procedure took, on average, 30 minutes to complete. Upon completion, each participant was paid $3.25.

Regression analysis. We attempted to predict each participant's inference about the liberalness of each of the 20 targets on the basis of the target's gender, age, attractiveness, ethnicity, cultural signifiers, and the

[2]Participants also indicated their own attitudes on the same statements used to assess their inferences about the targets' attitudes. Half the participants, randomly selected, indicated their own attitudes before inferring target attitudes. The order was reversed for the remaining participants. No order effects were observed. Participants' own attitudes are not directly relevant to the present study and, with the exception of the attitude scale data reported in Table 3, and one follow-up test, will not be referred to further.

extent to which the target was smiling. All predictors were entered simultaneously. A separate regression analysis was carried out on each participant's judgements about the liberalness of the 20 targets they saw (see Cooksey, 1996, for discussion of policy-capturing issues). We recognise that a regression analysis with eight predictors and only 20 cases is contrary to accepted practice in *inferential* analyses. In the present case, however, our purpose was simply to produce *descriptive* indices that could then be examined for consistency across independent participants. It should be noted as well that the criterion variable, the inference about target liberalness, was the combination of 10 separate judgements by the participant and thus was likely to be more reliable than any single measure.

Two descriptive indices were recorded for each participant: (1) the total R^2 indicating the proportion of variance accounted for by all eight predictors and (2) the zero-order correlation between each predictor and liberalness judgements, indicating the strength and direction of the relationship.

Results and discussion

Total R^2

Across the 60 participants, R^2 values ranged from .37 to .94. The mean R^2 value was .68, 95% $CI = .65$ to .72. These results show that multiple regression accounted for a substantial portion of the variability in the social inferences with the eight physical characteristics. In other words, regression was able to capture successfully individual participants' inferences about the attitude liberalness of others. This finding lends support to the idea that people are willing to use physical cues to make inferences about someone's attitudes. It is likely, however, that these results reflect a degree of overfitting. The need for cross-validation is clear and will be addressed later in this paper.

Zero-order correlations

The zero-order correlation between each predictor and liberalness judgements was calculated for each participant. Summary statistics for these correlations across the 60 participants are shown in Table 4. In accordance with the recommendations of Silver and Dunlap (1987), a Fisher r-to-z transformation was applied to the zero-order correlations before averaging and testing for significance. The mean of the transformed correlations differed significantly from zero for all but one of the predictors, *Caucasian*. There appears to have been considerable consensus across participants about the direction of the relationship between each cue and liberalness of attitudes. Participants tended to associate liberal attitudes most often with the following cues: female, young, attractive, Asian, Black, absence of a cultural signifier, and smiling.

TABLE 4
Zero-order correlations between predictors and judgements of target liberalism in Study 1

Predictor	Mean[b]	95% CI	t(59) (versus zero)	Percent of participants for whom row predictor ranked[a] 1 (smallest)	2	3	4	5	6	7	8 (largest)
Gender	0.35	0.29 to 0.41	11.38	5.3	12.3	12.3	21.1	14.0	5.3	19.3	10.5
Age	−0.45	−0.50 to −0.40	−15.51	3.5	0.0	8.8	8.8	15.8	35.1	19.3	8.8
Attractiveness	0.59	0.55 to 0.63	22.12	0.0	0.0	5.3	0.0	5.3	10.5	26.3	52.6
Asian	0.17	0.11 to 0.22	6.39	26.3	31.6	17.5	8.8	3.5	5.3	5.3	1.8
Black	0.24	0.19 to 0.29	8.88	21.1	12.3	17.5	15.8	14.0	14.0	5.3	0.0
Caucasian	−0.03	−0.11 to 0.05	−0.75[c]	24.6	21.1	17.5	10.5	10.5	5.3	5.3	5.3
Cultural signifier	−0.38	−0.43 to −0.32	−12.47[d]	10.5	7.0	7.0	15.8	17.5	17.5	15.8	8.8
Smiling	0.28	0.22 to 0.35	8.69	8.8	15.8	14.0	19.3	19.3	7.0	3.5	12.3

Three participants saw targets who did not vary on one of the predictors. Correlations in these cases could not be calculated.
[a]Rankings are based on absolute values of correlation coefficients. Ties were broken randomly. Percentages are based on 57 participants for whom data were complete. [b]Means and *CI* values are back-transformed *r* values following Fisher's *z* transformation (see text). [c]df = 58. [d]df = 57.

The single most important cue was target attractiveness. For 52.8% of participants this was the cue with the largest absolute r value. This finding is consistent with a wealth of previous research showing that attractiveness is a frequently used cue for making a wide variety of social inferences (see Eagly et al., 1991). In the present case it is possible that participants simply assumed that attractive targets would have attitudes similar to their own. To examine this possibility, we looked to see if the relationship between attractiveness and perceived target liberalness was weaker or possibly reversed among participants whose own attitudes were relatively conservative. The correlation between participants' own liberalness (see footnote 2) and their zero-order correlations between target attractiveness and liberalness, although in the expected direction, was not significant, $r(58) = .116$, $p = .379$. However, this correlation may have been attenuated by the restricted range of participants' attitude scores. Only 4 of our 60 participants (6.7%) had attitude scores on the conservative side of the scale midpoint.

As can be seen, the second strongest cue for making inferences about attitudinal liberalness was age, followed by the presence of a cultural signifier and gender, respectively. The age and gender influences are consistent with previous research on judgements about others' attitudes (see Grant et al., 1997, 2001). The influence of age may reflect the prevailing cultural view that older people are more committed to tradition and more resistant to change (e.g., Bassili & Reil, 1981; Truett, 1993). The influence of smiling may stem from the fact that liberal attitudes frequently reflect a more kindly position on many issues. This might also explain the tendency to associate liberal attitudes with the more nurturing characteristics often ascribed to women. To our knowledge, the relationship between attitudes and cultural signifiers has not been reported. It may be that the presence of a cultural signifier, at least to Western participants, is indicative of strong religious beliefs that are associated, in turn, with conservatism on other issues (Roccas, 2005). Some evidence for this speculation can be found in the present data. We examined, for each participant, the strength of the relationship between the presence of cultural signifiers and inferences on each of the five issues that make up the scale of liberalism. The relationship with religion (Mean $R^2 = .18$, 95% $CI = .15$ to .22) was the strongest of the five, followed by feminism (Mean $R^2 = .16$, 95% $CI = .12$ to .19), homosexuality (Mean $R^2 = .14$, 95% $CI = .11$ to .18), discipline (Mean $R^2 = .10$, 95% $CI = .07$ to .13), and immigration (Mean $R^2 = .09$, 95% $CI = .06$ to .11).

Cross validation

We attempted to assess the generalisability of the regression model for each participant by using two techniques to estimate a cross-validated R^2, one suggested by Rozeboom (1978) and a second method commonly referred to as a jack-knife procedure (Cooksey, 1996).

Rozeboom's R^2. For each participant an estimate of the cross-validated R^2 was calculated as follows:

$$Rozeboom's\ R^2 = 1 - \frac{(1 - R^2)(N + K)}{N - K}$$

In this equation, R^2 is the multiple squared correlation, N is the number of cases, and K is the number of predictors.[3] The more predictors there are in the model, relative to the number of cases, the more unreliable the model and the smaller the value of Rozeboom's R^2. In the present study, the mean Rozeboom's R^2, calculated across participants, was .29, 95% $CI = .22$ to .36.

Jack-knife R^2. In this procedure repeated analyses are conducted, each one on a sample of cases with a different single case deleted. The parameters of the model based on $N - 1$ cases are used to obtain a predicted value for the deleted case. We followed this procedure for each of our participants, obtaining a predicted liberalness score for each target using a regression equation based on the data for the other 19 targets. We then calculated the squared correlation between predicted and observed liberalness scores across the 20 targets for each participant. In order to avoid a positive bias, R^2 was set to zero for the small number of cases where the original correlation was negative. The resulting R^2 values had a mean of .24, 95% $CI = .18$ to .30.

Although we found that about two thirds of the variance in the 20 liberalness inferences was accounted for by the eight target cues, the reliability of the regression coefficients with such a high cue-to-cases ratio is a matter of concern. The Rozeboom and Jack-knife procedures addressed this concern and both procedures yielded similar estimates. These two calculations indicate that the original model overfitted the data as there was, on average, 39% and 44% drop in the percentage of variance accounted for by the models, respectively. Although the model does a good job of describing the relationship between physical cues and social inferences, the overfitting points to the need to test the model on new data. The Rozeboom and Jack-knife procedures are useful but they are not substitutes for testing the model on new datasets.

[3]The correction provided by Rozeboom's formula is similar in purpose and effect to the adjusted R^2 value provided by regression analyses in many statistical packages. Both procedures penalise overly complex models where overfitting is likely to be a problem. We chose Rozeboom's formula because it is the more conservative approach; it imposes a larger penalty, especially when the original R^2 value is relatively small.

In the second study we tried to enhance the generalisability of our results by having participants respond to a much larger set of target pictures. In this way we lessened the concern that our results might depend on the particular characteristics of a small set of cases and made it possible to estimate model parameters using data from the first half of the targets that each participant saw and test the models using data from the last half. Finally, we extended the previous study by examining a different domain of social judgement. In this case, participants were asked to make inferences about intelligence.

STUDY 1

Method

Participants

Participants were 40 undergraduates (20 men and 20 women) at Memorial University.

Procedure

The procedure used in Study 2 was followed, except that each participant saw pictures of all 288 targets. The order of presentation was randomised for each participant. As each picture appeared, the participant was asked to rate the intelligence of the person shown on a scale from 1 (not at all intelligent) to 7 (very intelligent).

A separate regression analysis was carried out on each participant's judgements about the intelligence of the first half (144) of the targets they saw. In this analysis all eight predictors described earlier were entered together. As in the first study, we calculated the total R^2 for all eight predictors together and the zero-order correlation between each predictor and intelligence judgements.

Results and discussion

Total R^2

Across the 40 participants, R^2 values ranged from .04 to .55. The mean value was .21, 95% $CI = .18$ to .25. When the Rozeboom correction described earlier was applied, the mean R^2 was reduced to .16, 95% $CI = .13$ to .20. In this inference domain, multiple regression appears to account for a relatively small portion of the variability in social inferences. In comparison to attitudinal inferences, where the mean Rozeboom R^2 value was .29, this finding suggests that people are less willing to use physical cues to make inferences about someone's intellectual abilities.

Zero-order correlations

The zero-order correlation between each predictor and intelligence judgements was calculated for each participant. Summary statistics for these correlations across the 40 participants are shown in Table 5. The mean correlations differed significantly from zero for six of the eight predictors; target gender and age were not statistically significant. Higher ratings of intelligence were most often associated with the following cues: attractiveness, Asian, Caucasian, and smiling. Lower ratings were associated with displaying a cultural signifier and being Black.

The single most important cue was whether or not the target was Asian. For 32.5% of participants, Asian was the cue with the largest absolute r value. The fact that Asians were rated relatively highly on intelligence is consistent with the competence stereotype reported by Lin, Kwan, Cheung, and Fiske (2005). The next strongest cue, attractiveness, had the highest absolute r value for 22.5% of participants, a finding that is consistent with Eagly et al.'s (1991) results. Intelligence was also related positively to smiling and negatively to cultural signifiers. We suspect that the presence of a smile triggers a general positivity bias that enhances favourability of ratings. As yet, we can only speculate about the role of cultural signifiers in judgements of intelligence. It may be that perceptions of limited educational and economic opportunities, negativity bias, and ethnocentrism all play a role in the impact of this particular physical cue.

Cross validation

We assessed the generalisability of the regression model for each participant by using the regression equation derived from that participant's ratings of the first half of the targets to predict his or her ratings for the last half of the targets. One female participant who assigned identical ratings to all the targets in the last set was eliminated from this analysis. Across the other 39 participants, the mean squared correlation between predicted and actual ratings on the last set of targets was .14, 95% $CI = .11$ to .17. As expected, the regression model did less well on the test data set than on the training data set where, as reported above, the mean R^2 was .21.

The problem of overfitting increases with the complexity of the model, that is, with the number of parameters that are fitted. We compared the results reported above with those for a simple one-cue model. For each participant we selected the single cue that correlated most strongly with judgements of intelligence on the first half of the targets. We then used this

TABLE 5
Zero-order correlations between predictors and judgements of target intelligence in Study 2

				Percent of participants for whom row predictor ranked[a]							
Predictor	*Mean*[b]	*95% CI*	*t(39) (versus zero)*	*1 (smallest)*	*2*	*3*	*4*	*5*	*6*	*7*	*8 (largest)*
Gender	−0.04	−.09 to .01	−1.79	15.0	15.0	10.0	15.0	17.5	7.5	10.0	10.0
Age	−0.04	−.10 to .01	−1.64	20.0	22.5	10.0	10.0	7.5	12.5	12.5	5.0
Attractiveness	0.18	.12 to .23	5.92	5.0	15.0	12.5	7.5	2.5	17.5	17.5	22.5
Asian	0.19	.15 to .23	9.64	5.0	7.5	2.5	7.5	12.5	20.0	12.5	32.5
Black	−0.05	−.09 to −.01	−2.51	12.5	17.5	15.0	22.5	12.5	5.0	7.5	7.5
Caucasian	0.07	.03 to .10	3.74	17.5	10.0	22.5	15.0	12.5	15.0	5.0	2.5
Cultural Signifier	−0.17	−.23 to −.11	−5.52	5.0	2.5	5.0	10.0	15.0	17.5	27.5	17.5
Smiling	0.08	.05 to .12	4.61	20.0	10.0	22.5	12.5	20.0	5.0	7.5	2.5

[a]Rankings are based on absolute values of correlation coefficients. Ties were broken randomly. Percentages are based on all 40 participants. [b]Means and *CI* values are back-transformed *r* values following Fisher's *z* transformation (see text).

cue to predict judgements on the first and second set of targets. The mean R^2 values for this single-cue model were .11 when applied to the first set of targets (95% $CI = .08$ to .13) and .09 when applied to the second set (95% $CI = .06$ to .12). Although overfitting is less of a problem with this single-cue model, it is significantly less accurate than the multiple-cue model in predicting the new data, $t(38) = 5.16$, $p < .001$. The superior performance of the multiple-cue model contrasts with recent findings (e.g., Dhami, 2003; Gigerenzer et al., 1999) that single-cue models do as well as more complex ones on cross validation data. Further research is needed to determine whether the superiority of the more complex model in the present case is specific to intelligence judgements or is characteristic of social inferences in general. As illustrated by Czerlinski, Gigerenzer, and Goldstein (1999), it is important that comparisons of the sort made here are conducted on cross validation data in order to provide a fair comparison of the predictive utility of models and to determine the conditions under which simple heuristics succeed and fail.

We should note, however, that the R^2 values for the single-cue model may be low because the model does not capture the possibility that cue usage can sometimes be target specific. In other words, people may use different cues for different pictures, especially those that are salient in the picture. For example, to judge intelligence, a participant may use age for an extremely old target, cultural garb for someone wearing traditional Innu dress, or attractiveness for a particularly good-looking person.

Idiosyncratic stereotypes

From the analyses of the zero-order correlations it is clear that there is a large consensual component to people's stereotypes about the particular target cues we studied. But is it the case that all participants weighted these cues in the same way? We addressed this question by again examining each participant's judgements about the last half of the targets. In this case we compared predictions based on that participant's own equation, estimated from his or her inferences on the first half of the targets, with predictions based on the equation of a different participant. For purposes of this comparison, each participant was paired randomly with a different "partner". Thus each participant's equation was used twice: once to predict his or her own later inferences and once to predict the inferences of another participant.

If a different participant's equation is as good as one's own, this indicates that the consensual stereotype is paramount and that little is to be gained by considering idiosyncratic variation. But if a different participant's equation is worse than one's own, this can be taken as evidence of individual differences in the way the cues are used. As reported above, predictions based on the participant's own equation yielded an R^2 value of .14. The corresponding figure for predictions based on a different participant's equation, $R^2 = .07$, 95% $CI = .05$ to .10, was significantly lower, $t(38) = 4.60$, $p < .001$. This "random partner" method of assessing idiosyncratic cue usage is a useful method for differentiating between consensual and idiosyncratic components of judgements and should be considered in future studies of this kind.

CONCLUDING THOUGHTS

The purpose of the present paper was to learn about the relationships between physical characteristics and people's inferences about the attitudes and intelligence of others. Consistent with previous research, we found that people tend to rely on the same readily available physical cues, especially attractiveness, to make social inferences. Attractiveness, like Asch's (1946) "warm–cold" variable, may act as a central trait that captures the information associated with a number of other cues. As with other kinds of decisions (e.g., Dhami, 2003; Smith & Gilhooly, 2006), perceivers seem to have little difficulty making snap judgements about how liberal or intelligent others are, judgements that often have many important social consequences (e.g., voting for a political candidate or hiring a job applicant).

It is important to note that, although the regression models inevitably showed some degree of overfitting, they continued to have predictive value when, in Study 2, they were applied to a new set of data. It appears that the stereotypic relationships between physical cues and social inferences are reasonably stable. At the same time, however, our results indicate that there are notable individual differences in the way people make social inferences. Further research is needed to determine how the weights that people give to the various cues are related to the perceivers' other characteristics (e.g., their age and gender) as well as to the domain and circumstances in which the judgements are made.

Our results indicate that, across targets, many cues are used in the kinds of social judgements studied here. Unlike previous research showing similar performance of multiple regression and simple heuristics models on cross-validation (see Czerlinski et al., 1999; Dhami, 2003), we found that multiple regression models were better able to capture intelligence judgements than single-cue models. In contrast to two-alternative, forced-choice decisions, it may be that social inferences are influenced by more cues and that the multiple regression models are more appropriate for capturing judgement policies in this domain. However, it could also be argued that the superior performance of the compensatory multiple-cue models is a function of the nature of multiple regression analysis, which provides a best-fitting summary of cue use across all targets. Therefore we cannot conclude that people integrate multiple cues in their judgements of any *single* target. One possibility is that people might use different cues for different pictures, and that they do so in a non-compensatory way. If this is the case, it may well improve the ability of a single-cue model to capture participant judgements. It may be that a heuristic based on just a single *salient* cue (e.g., one that is extreme or stands out in a particular context), drives many social judgements. For example, a strikingly attractive target may be judged to be liberal regardless of the presence of other cues (e.g., old, male, traditional dress) that would point in the other direction. In our view, determining what prompts the use of a particular cue (e.g., salience) is essential if were are to capture the nature of social judgements, test the validity of various models, and resolve some of the issues raised in the current research.

Manuscript received 9 April 2008
Revised manuscript received 5 June 2009
First published online 22 August 2009

REFERENCES

Abelson, R. P., & Levi, A. (1985). Decision making and decision theory. In G. Lindzey & E. Arsonson (Eds.), *Handbook of social psychology* (Vol. 2, pp. 231–309). New York: Random House.

Alley, T. R. (1988). *Social and applied aspects of perceiving faces*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Anderson, N. H. (1982). *Methods of information integration theory*. New York: Academic Press.

Asch, S. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, *41*, 258–290.

Bassili, J. N., & Reil, J. E. (1981). On the dominance of the old-age stereotype. *Journal of Gerontology*, *36*, 682–688.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Skrull & R. S. Wyer (Eds.), *Advances in social cognition* (pp. 1–36). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Brunswick, E. (1945). Social perception of traits from photographs. *Psychological Bulletin*, *42*, 535–536.

Bull, R., & Rumsey, N. (1988). *The social psychology of facial appearance*. New York: Springer-Verlag.

Cooksey, R. W. (1996). *Judgement analysis: Theory, methods, and applications*. New York: Academic Press.

Czerlinski, J., Gigerenzer, G., & Goldstein D. G. (1999). In G. Gigerenzer, P. M. Todd, & The ABC Research Group, *Simple heuristics that make us smart* (pp. 97–118). New York: Oxford University Press.

Dalgleish, L. I. (1988). Decision making in child abuse cases: Applications of social judgement theory and signal detection theory. In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgement: The SJT view* (pp. 317–360). Amsterdam: Elsevier.

Dhami, M. K. (2003). Psychological models of professional decision making. *Psychological Science*, *14*, 175–180.

Dhami, M. K., & Harries, C. (2001). Fast and frugal versus regression models of human judgement. *Thinking and Reasoning*, *7*, 5–27.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*, 285–290.

Doherty, M. E., & Brehmer, B. (1997). The paramorphic representation of clinical judgement: A 30 year retrospective. In W. M. Goldstein & R. M. Hogarth (Eds.), *Research on judgement and decision making: Currents, connections, and controversies*. New York: Cambridge University Press.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but . . . : A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*, 109–128.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Toronto: Harcourt Brace Jovanovich College Publishers.

Efron, B. (1982). *The jack-knife, the bootstrap, and other resampling plans*. Philadelphia, PA: Society for Industrial Mathematics.

Fishbein, M. (1980). A theory of reasoned action: Some applications and implications. In H. E. Howe Jr., & M. M. Page (Eds.), *Nebraska Symposium on Motivation,* 1979 (Vol. 27, pp. 65–116). Lincoln, NE: University of Nebraska Press.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–74). San Diego, CA: Academic Press.

Gigerenzer, G., Todd, P. M., & The ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Grant, M. J., Button, C. M., Hannah, T. E., & Ross, A. S. (2000). *Perceptions of age and attractiveness by men and women of different ages*. Poster presented at the meeting of the XXVII International Congress of Psychology, Stockholm, Sweden.

Grant, M. J., Button, C. M., Hannah, T. E., & Ross, A. S. (2002). Uncovering the multidimensional nature of stereotype inferences: A within-participants study of gender, age, and physical attractiveness. *Current Research in Social Psychology*, *8*, 19–38.

Grant, M. J., Button, C. M., Ross, A. S., & Hannah, T. E. (1997). Accuracy of attitude stereotypes: The case of inferences based on gender. *Canadian Journal of Behavioural Science*, *29*, 83–91.

Grant, M. J., Ross, A. S., Button, C. M., Hannah, T. E., & Hoskins, R. (2001). Attitudes and stereotypes about attitudes across the lifespan. *Social Behavior & Personality*, *29*, 749–762.

Hammond, K. R. (1996). *Human judgement and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. New York: Oxford University Press.

Hatfield, E., & Sprecher, S. (1986). *Mirror, mirror: The importance of looks in everyday life*. New York: SUNY Press.

Heald, J. E. (1991). Social judgement theory: Applications to educational decision making. *Educational Administration Quarterly*, *27*, 343–357.

Herman, C. P., Zanna, M. P., & Higgins, E. T. (1986). *Physical appearance, stigma, and social behaviour: The Ontario Symposium* (Vol. 3). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Himmelfarb, S., & Anderson, N. H. (1975). Integration theory applied to opinion attribution. *Journal of Personality and Social Psychology*, *31*, 1064–1072.

Hoffman, P. J. (1960). The paramorphic representation of clinical judgement. *Psychological Bulletin*, *57*, 116–131.

Lin, M. H., Kwan, V. S. Y., Cheung, A., & Fiske, S. T. (2005). Stereotype content model explains prejudice for an envied outgroup: Scale of anti-Asian American stereotypes. *Personality and Social Psychology Bulletin*, *31*, 34–47.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*, 93–120.

Rhodes, G., & Zebrowitz, L. A. (2002). *Facial attractiveness: Evolutionary, cognitive, and social perspectives*. Norwood, NJ: Ablex.

Roccas, S. (2005). Religion and value systems. *Journal of Social Issues*, *61*, 747–759.

Rozeboom, W. W. (1978). Estimation of cross-validated multiple correlation: A clarification. *Psychological Bulletin*, *85*, 1348–1351.

Silver, N. C., & Dunlap, W. P. (1987). Averaging correlation coefficients: Should Fisher's z transformation be used? *Journal of Applied Psychology*, *72*, 146–148.

Smith, L., & Gilhooly, K. (2006). Regression versus fast and frugal models of decision-making: The case of prescribing for depression. *Applied Cognitive Psychology*, *20*, 265–274.

Truett, K. R. (1993). Age differences in conservatism. *Personality and Individual Differences*, *14*, 405–411.

Wyer, R. S. Jr. (1974). *Cognitive organization and change: An information-processing approach*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.