Individual differences in judgmental tendencies derived from first impressions

Alan R. King*, Allison N. Pate

Psychology Department, University of North Dakota, PO Box 8380, Grand Forks, ND 58202-8380, USA

Abstract

This report introduces the First Impression Interaction Procedure (FIIP) as a new method of assessing participant initial reactions to unfamiliar partners encountered during 25-min uncontrolled interaction sessions. The 14-item FIIP rating form quantifies first impression central tendency (Judgment Index), variability (Judgment Variability Index), and the degree to which participants judged others more harshly than self (Judgment Ratio). Selected participants can also be distinguished by their tendency to form and express unusually harsh first impressions of others. College students ($n = 252$) completed two FIIP interaction sessions with six partners. Internal consistency and temporal stability coefficients for the three FIIP measures ranged from 0.85 to 0.94. Judgment Index scores in excess of 1 S.D. from the mean were shown to reflect a generalized tendency for participants to form and express first impressions that were typically 1.5–2 S.D.s harsher than 80% of their interaction partners. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: First impressions; First impression interaction procedure (FIIP); Judgmental tendencies; Judgmental trait; Judgmental personality

Many different situation and person variables have been found to effect the formation of first impressions. Factors such as physical attractiveness (Cash, Gillen, & Burns, 1977), race (Hart & Morry, 1997), voice quality (Berry, Hansen, Landry-Pester, & Meier, 1994), manner of dress (Temple & Loewen, 1993), perceived material wealth (Dittmar, 1992), similarity (Lydon, Jamieson, & Zanna, 1988), expressive style (Costanzo & Archer, 1989), nonverbal behavior (Baum, Fisher, & Singer, 1985), first names (Steele & Smithwick, 1989), and many others have been shown to affect the first impressions of the average observer.

---

* Corresponding author. Tel.: +1-701-777-3451.
 E-mail address: alan_king@und.nodak.edu (A.R. King).

Social judgment (Sherif & Hovland, 1961) and accentuation theories (Eiser & Stroebe, 1972) have provided general models which emphasize the broad and subjective categorizations such as acceptability versus unacceptability which influence initial impression formation. Judgment similarities are often accentuated among members of the same group (Doise, Deschamps, & Meyer, 1978; Taifel, 1978), while differences are magnified in social comparisons of targets viewed as differing in some important categorical feature (Taifel & Wilkes, 1963).

Contemporary social identity (Tajfel & Turner, 1986) and related self-categorization (Turner, Hogg, Oakes, Reicher, & Wetherall, 1987) theory notes the importance of self-perceptions in the formation of impressions about the qualities of others. The valence of social impressions are thought to be accentuated by the degree to which participants conceptually group targets into categories that are associated with personal favorable or unfavorable expectancies. Social judgments of ingroup members do tend to be more favorable than the impressions formed of others not sharing the same group membership (Mullen, Brown, & Smith, 1992), but these biases can be attenuated by exposure to credible opposing views or information (Doosje, Spears, & Koomen, 1995; Spears & Manstead, 1989). Social judgments may also become less variable among ingroup members in contrast to the impressions formed of additional targets conceptually classified into other outgroups (Doosje, Haslam, Spears, Oakes, & Koomen, 1998). Social judgments may also be influenced by target conformity or violation of stereotypic gender expectancies (Birdwhistell, 1970).

First impressions appear to serve an important role in initiation and maintenance of social relations. Once formed, first impressions have been shown to exhibit considerable resistance to change (Asch, 1946, 1952; DiGirolamo & Hintzman, 1997; Dougherty, Turban, & Callender, 1994; Jones, Rock, Shaver, Goethals, & Ward, 1968). The stability of first impressions appears related to the temporal pattern of exposure and affective sequellae associated with their formation (Aronson & Linder, 1965; Mette, 1971). Generalized tendencies to routinely form and express harsh first impressions could be described as manifestations of a judgmental personality trait.

A conventional method for measuring first impression tendencies has yet to be established. Most first impression studies (over 150 now available) examine participant reactions to a small number of unique, precisely defined targets which are presented indirectly in the form of written vignettes, photographs, or videotaped segments. The direction (positive versus negative) and intensity of first impression ratings have been examined, but qualitative analyses of the adjectives used to describe targets have also been popular. Ratings scales in these studies have tended to be relatively simple and idiosyncratic to purposes of the respective investigation (Bryan, Coleman, Ganong, & Bryan, 1986; Walther, 1993). An alternative approach described in this report as the First Impression Interaction Procedure (FIIP) is distinguished by its reliance on diverse targets encountered in a more natural social setting as a strategy for generating ecologically valid measures of first impression central tendency, variability, and proclivity to judge others more or less harshly than self. FIIP estimates of judgment standards have the benefit of being derived from samples of actual rating behavior exhibited after direct target encounters.

The present investigation should be distinguished from other attempts to measure the accuracy of impressions via comparison with criterion variables that represent some sense of "reality" or "truth" regarding the personal qualities under inspection. Cronbach's (1955) and others have warned of the added methodologic challenges inherent in that endeavor (Costanzo & Archer, 1989; Frey & Smith, 1993; Kenny & Albright, 1987; Kenny & DePaulo, 1993; Levesque & Kenny, 1993). The two main objectives pursued in this report were modest by comparison. First, an effort

was made to provide an analysis of the psychometric soundness of this new quasi- naturalistic approach to quantify generalized impression formation tendencies. Second, an attempt was made to identify individuals who were distinguished by their tendency to form and express unusually harsh first impressions of others. This search for a judgmental personality trait was consistent with Cronbach's (1955) historic conclusion that impression formation was often determined more by the judge than the interaction.

1. Method

1.1. Participants

Undergraduate students (approximately 50% of those invited) enrolled in selected psychology courses at the University of North Dakota agreed to participate in a study of interpersonal perceptiveness requiring 4 h of time to earn about 4% of extra course credit. Data was analyzed for participants completing two separate interaction sessions with a total of 6 unfamiliar partners. A total of 184 women and 68 men comprised the final samples after exclusions due to missed appointments or familiarity with one or more of the assigned partners. Familiarity was assessed for each partner using a five-point scale which ranged from 0 (never met or formed any impression) to 4 (know person well and have stable impressions). A score of two (have had significant contact, but not formed any firm impressions) for any of the six partners resulted in exclusion from the study. The 252 final participants were primarily white (92%) or Native American (about 8%), between the ages of 18 and 52 ($M = 24$, S.D. $= 7$), and primarily classified at a sophomore or junior academic level in terms of cumulative college credits ($M = 70$, S.D. $= 30$). Two additional independent samples selected from the same pool were used to further establish test–retest reliability ($n = 18$) and concurrent validity ($n = 41$).

2. Materials

Several graduate assistants participated in a series of meetings over an academic semester to consider the merits of a list of descriptive terms used by prior impression formation researchers. The objective was to identify 10 to 15 descriptive dimensions that could offer a broad canvas of personal attributes that was internally consistent, uncomplicated in factor structure, and easy for users to conceptualize and apply. Prior research has suggested that an overall good versus bad evaluative factor strongly influences impressions formation (Osgood, Suci, & Tannenbaum, 1957; Rosenberg, & Olshan, 1970), and it was assumed that a representative sample of general personality qualities would meet the purposes of the study. Kenny, Albright, Malloy, and Kashy (1994) found no less than 407 trait descriptors applied in the 32 consensus accuracy studies reviewed in their important meta- analysis. Many of these items were considered for inclusion. Items with considerable overlap were merged into single dimensions, and those which required higher levels of inference or interpretability were eliminated. A dimensional rating scale was preferred since it would likely generalize to those commonly used in job interviews or other evaluative processes. Some initial pilot testing was conducted, and the scale construction process culminated in the 14

item FIIP rating form (Appendix A). Participants were also provided a laminated copy of the 14 attributes anchored on a seven-point rating scale (positive poll assigned score of 1; negative poll assigned score of 7) to assure understanding of how scores should be rendered.

2.1. Scoring

The FIIP rating scale used in the present study provided a basis for the calculation of a number of different indices of first impression formation. A Judgment Index (JI) can be derived from the average judgment rating across all 14 items and six partners. A Prediction Index (PI) can be calculated similarly as an indicator of the degree to which participants anticipate favorable or unfavorable ratings from others. A Self Index (SI) can also be derived from the self image ratings averaged over both sessions for the 14 FIIP items (self ratings not differentiated by individual partner). JI, PI and SI scores all range from 1 to 7 and can be calculated for either individual or some subset of the six partners.

As Cronbach observed in 1955, the meaning of dimensional ratings can, and typically do, vary as a function of the personal standards applied by the judge. The assignment of academic course grades offers a good example. One professor rarely assigns "A" grades and considers the "B" student to be exceptional. The same "B" grade given by another college professor might reflect a different judgment that the student's performance is relatively weak compared with the remainder of the class. The ratings themselves do not adequately reflect the judgment or attitude of the observer. The FIIP SI and PI ratings provide additional information which can be used to enhance the meaning of JI ratings. A JI score of 2.5 for a male participant appears to represent an unremarkable impression that does not seem to reflect a judgmental attitude. SI and PI scores of 1.0 for the same participant, however, would indicate that the participant sees himself, and expects to be viewed, in a far more favorable light than he is prepared to extend to others. Consideration was given to several ratios that could be used to quantify the discrepancy between impressions of self and others. PI/JI and SI/JI ratios were natural candidates, but preliminary reliability analyses suggested that a combination of these two ratios would promise a more stable index of the desired ratio. The Judgment Ratio (JR) was ultimately calculated as the square root of SI times PI divided by JI: $(\sqrt{SI \times PI})/JI$. The JR index ranges from 0.14 to 7.0, and scores below 1.0 indicate a more favorable perception of self than others.

The Judgment Variability Index (JVI) was calculated as the standard deviation of the JI scores generated for the partners. The JVI represents the degree to which participants discriminated in their application of JI ratings. JVI scores can range from 0 to 3.0 and provide an estimate of the degree to which situational factors such as partner attributes, mood state changes, and other factors influence the central tendency of participant first impressions. Low JVI scores suggest inclinations to judge others similarly despite the many differences which are certain to occur in partner behavior, conversational content, and other circumstances that may vary greatly from session to session. High JVI scores suggest responsiveness to partner attributes and interaction events that may be additional evidence of judgmental attitude. For example, one participant in the present study generated a JVI score of 0.55 (about average for the total distribution) and provided individual JI scores of 1.00, 1.07, 1.21, 1.36, 1.00 and 1.00. Another participant with a JVI score of 1.28 showed more variable ratings of 1.21, 1.36, 1.43, 2.79, 3.07 and 4.43. JVI scores as high as 1.28 require a range of individual JI scores which assure that one, or usually more, of

the six partners were being rated quite harshly, relative to the remainder of the group. Particularly high JVI scores suggest a penchant to form extreme, probably bipolar, evaluations of others.

2.2. Design and procedure

Participants were assigned to two separate interaction sessions with three partners in each group. The sessions were scheduled at least 1 week apart, and participants were provided the following expectation: "You will be assigned to meet in a particular room at a particular time with three partners. You will be instructed to simply interact freely with these partners for a 25 minute period which will hopefully provide an opportunity to get to know one another. At the conclusion of each interactive session, you will be provided a confidential space where you will be asked to provide your impressions, through various formats, of the personality attributes exhibited by each of your three partners. You also will be asked to make predictions about the manner in which each of your interaction partners rated you at the end of the preceding session."

Up to 36 participants began each session in a large classroom, followed by assignment of groups to individual rooms with color coded (by group) identification badges labeled A through D. Participants were informed that there were no restrictions on the conversation content and that it was acceptable to share their name or other personal information with the group. A matching (by letter) color-coded folder with the FIIP rating forms was distributed at the conclusion of the 25-min session to each group member who was ushered back into the large classroom and seated apart from their partners. Reusable laminated copies of the instructions and the 14-item FIIP rating scale assisted participant efforts to provide judgments, predictions and self ratings. FIIP rating forms were monitored when returned to assure proper completion.

A four-item exit questionnaire followed each session to assess participant reactions to the FIIP itself (e.g. seemed reasonable, relevant, interesting, worthy of recommendation, etc.) on a similar seven-point scale. General comments were solicited from an open-ended final question.

3. Results

3.1. JI

FIIP JI scores ranged from 1 to 4 or 5 for all items. Men provided significantly harsher ratings than women for all but four traits (good versus bad listener, interesting versus boring, authentic versus inauthentic, and self-assured versus self-conscious). Table 1 presents descriptive statistics for male and female JI scores as a function of session number and gender configuration (number of same and opposite sex partners). Male participants generated significantly harsher judgments for both sessions one, $F(1,244) = 6.69$, $P = 0.01$, and two, $F(1,244) = 10.8$, $P = 0.001$. Combined six-partner JI scores were also higher for male participants, $F(1,240) = 12.39$, $P < 0.001$. Male partners also received harsher ratings from both male, $F(1,22) = 6.15$, $P = 0.02$, and female, $F(1,46) = 6.64$, $P = 0.01$, participants. High correlations were still found between JI scores for men and women partners generated by male, $r(21) = 0.80$, and female, $r(45) = 0.81$, participants.

Table 1 does also show a significant first session partner gender configuration effect, $F(3,244) = 4.00$, $P = 0.008$. A Student–Newman–Keuls multiple range comparison indicated that

participants assigned to balanced first session configurations (two men and two women) elicited the harsher JI scores than those assigned to sessions of either all male or all female participants. Participant gender by partner gender interaction analysis indicated that this general configuration effect was similar for men and women participants, $F (3,244) = 0.24$, $P = 0.87$. This same configuration effect was not observed for second session JI scores, $F (3,244) = 0.67$, $P = 0.57$, and the gender and configuration interaction was similarly not significant for second session scores, $F (3,244) = 0.42$, $P = 0.74$.

Retest differences in JI scores were not found, $F (1,250) = 1.27$, $P = 0.26$, and the gender by retest interaction was similarly not significant, $F (1,250) = 1.24$, $P = 0.26$. Of most importance, the effects of configuration, $F (6,240) = .84$, $P = 0.54$, and the participant gender by configuration interaction, $F (4, 240) = 0.49$, $P = 0.74$, were not significant for six partner analyses. While balanced first session assignments might be avoided, the usage of data generated from six partners over two sessions appears to minimize potential gender partner effects on JI scores.

3.2. PI

Men predicted significantly harsher evaluations than women for both sessions one, $F (1,244) = 9.91$, $P = 0.002$, and two, $F (1,244) = 11.40$, $P < 0.001$. Combined six-partner PI scores were also higher among male ($M = 2.62$, S.D. $= 0.78$) than female ($M = 2.27$, S.D. $= 0.61$) participants, F

Table 1
Gender analyses of Judgment Index (JI) scores

Configuration	Male			Female			Male			Female		
	M	S.D.	n	M	S.D.	n	M	S.D.	n	M	S.D.	n
	First Session JI Scores						Second Session JI Scores					
Three Partners	2.52	0.74	68	2.27	0.70	184	2.52	0.99	68	2.13	0.78	184
0 Same/3 Opp	2.26	75	21	2.57	0	1	2.49	0.97	19	2.47	1.02	2
1 Same/2 Opp	2.69	0.75	35	2.54	0.72	25	2.54	1.10	38	2.25	0.76	26
2 Same/1 Opp	2.59	0.77	5	2.36	0.81	49	2.91	0.39	4	2.16	0.70	60
3 Same/0 Opp	2.45	0.49	7	2.15	0.63	109	2.30	0.66	7	2.07	0.83	96
				Two session combined JI scores								
Four Partners				2.04	0.44	8	2.69	0.29	3			
Five Partners				2.14	0.59	77	2.69	0.79	31			
Six Partners				2.20	0.62	184	2.52	0.72	68			
0 Same/6 Opp				—	—	—	2.30	0.84	12			
1 Same/5 Opp				—	—	—	2.51	0.70	11			
2 Same/4 Opp				2.27	0.50	8	2.61	0.74	35			
3 Same/3 Opp				2.44	0.62	20	2.98	0.00	1			
4 Same/2 Opp				2.21	0.69	41	2.88	0.61	2			
5 Same/1 Opp				2.19	0.57	46	1.76	0.00	1			
6 Same/0 Opp				2.12	0.63	69	2.39	0.57	6			

Men rendered harsher ratings ($P < 0.05$) for the first, second and combined session analyses. Six partner configuration or gender interaction effects were not observed, but participants in the first session with two opposite gender partners did score higher than those with 0 or 3 ($P < 0.05$).

$(1,240) = 13.86$, $P < 0.001$. Configuration and gender by configuration effects were not observed for any (first session, second session, combined six-partner) analyses. PI scores did decrease significantly from session one to two, $F(1,250) = 4.69$, $P = 0.03$, but the gender by retest interaction was not significant, $F(1,250) = 0.17$, $P = 0.68$.

3.3. SI

SI scores generated by women participants were more favorable than men for both the first, $F(1,234) = 9.83$, $P = 0.002$, and second sessions, $F(1,234) = 3.86$, $P = 0.05$. Combined six-partner SI scores were also lower among female ($M = 2.09$, S.D. $= 0.56$) than male ($M = 2.40$, S.D. $= 0.70$) participants, $F(1,230) = 12.31$, $P < 0.001$. Configuration and gender by configuration effects were not observed for any (first session, second session, combined six-partner) analyses. SI scores were not found to change significantly from session one to two, $F(1,240) = 3.58$, $P = 0.06$, and the gender by retest interaction was also not significant, $F(1,240) = 0.01$, $P = 0.96$.

3.4. JR

Table 2 indicates that participant gender differences were not found for the first, second, or combined six-partner JR analyses. Configuration and gender by configuration effects were similarly not observed. JR scores were not found to change significantly from session one to two, F

Table 2
Gender analyses of Judgment Ratio scores $(\sqrt{SI \times PI}))/JI$

Configuration	Male			Female			Male			Female		
	M	S.D.	*n*	*M*	S.D.	*n*	*M*	S.D.	*n*	*M*	S.D.	*n*
	First Session JR Scores						Second Session JR Scores					
Three Partners	*1.03*	*0.27*	*67*	*1.02*	*0.24*	*175*	*1.02*	*0.26*	*67*	*1.04*	*0.26*	*175*
0 Same/3 Opp	1.12	0.36	21	1.06	0.00	1	0.99	0.21	18	0.99	0.24	2
1 Same/2 Opp 2.69	0.98	0.19	35	0.99	0.19	22	1.03	0.30	38	1.03	0.23	25
2 Same/1 Opp 2.59	0.99	0.34	4	1.01	0.24	48	1.03	0.09	4	1.03	0.28	57
3 Same/0 Opp	1.02	0.24	7	1.04	0.24	104	0.98	0.23	7	0.98	0.37	91
				Two session combined JR scores								
Four Partners				1.04	0.21	8	0.95	0.10	3			
Five Partners				1.04	0.21	73	0.97	0.17	31			
Six Partners				*1.02*	*0.22*	*175*	*1.01*	*0.20*	*67*			
0 Same/6 Opp				—	—	—	1.00	0.18	12			
1 Same/5 Opp				—	—	—	1.09	0.34	11			
2 Same/4 Opp				1.02	0.19	7	0.99	0.13	34			
3 Same/3 Opp				0.95	0.20	19	1.06	0.00	1			
4 Same/2 Opp				1.04	0.23	40	1.00	0.27	2			
5 Same/1 Opp				0.98	0.22	43	1.47	0.00	1			
6 Same/0 Opp				1.05	0.21	66	0.93	0.10	6			

Gender and gender by configuration differences were not significant for the first, second and combined session analyses. Retest differences were not significant.

$(1,240) = 0.07$, $P = 0.79$, and the gender by retest interaction was also not significant, F $(1,240) = 0.45$, $P = 0.50$.

3.5. JVI

Table 3 shows that participant gender differences were not found for the first, second, or combined six-partner JVI analyses. Configuration and gender by configuration effects were not observed. JVI scores were found to diminish significantly from session one to two, $F(1,250) = 4.35$, $P = 0.04$, but the gender by retest interaction was not significant, $F(1,250) = 0.01$, $P = 0.94$.

3.6. Scale development

3.6.1. Factor structure

A principal component analysis demonstrated that one primary factor in six-partner JI scores would account for 57% of the total variance using Kaiser's (1960) retention criterion (eigenvalue > 1). Men and women showed similar factor structure with 80% agreement regarding the composition of the top and bottom five item loading strengths. Principal component factor structures for PI and SI scores were somewhat more complicated. Most of the variance in PI and SI scores (54 and 45%, respectively) was similarly accounted for by single factors. Stevens' (1996) recommended

Table 3
Gender analyses of Judgment Variability Index (JVI) scores

Configuration	Male			Female			Male			Female		
	M	S.D.	n	M	S.D.	n	M	S.D.	n	M	S.D.	n
	First Session JVI scores						Second Session JVI scores					
Three Partners	*0.41*	*0.36*	*68*	*0.42*	*0.43*	*184*	*0.33*	*0.38*	*68*	*0.35*	*0.45*	*184*
0 Same/3 Opp	0.35	0.36	21	0.31	0	1	0.23	0.22	19	0.08	0.05	2
1 Same/2 Opp 2.69	0.46	0.39	35	0.43	0.35	25	0.40	0.45	38	0.38	0.41	26
2 Same/1 Opp 2.59	0.54	0.36	5	0.50	0.53	49	0.42	0.33	4	0.41	0.50	60
3 Same/0 Opp	0.28	0.15	7	0.39	0.39	109	0.19	0.19	7	0.31	0.43	96
				Two session combined JVI scores								
Four Partners				0.66	0.46	8	0.49	0.36	3			
Five Partners				0.58	0.45	77	0.54	0.24	31			
Six Partners				*0.55*	*0.39*	*184*	*0.58*	*0.40*	*68*			
0 Same/6 Opp				—	—	—	0.44	0.24	12			
1 Same/5 Opp				—	—	—	0.55	0.28	11			
2 Same/4 Opp				0.47	0.33	8	0.70	0.49	35			
3 Same/3 Opp				0.60	0.47	20	0.19	0.00	1			
4 Same/2 Opp				0.61	0.42	41	0.67	0.24	2			
5 Same/1 Opp				0.58	0.35	46	0.35	0.00	1			
6 Same/0 Opp				0.48	0.37	69	0.33	0.11	6			

Gender and gender by configuration differences were not significant for the first, second and combined session analyses. JVI scores decreased significantly from session one to two (retest effect) for both male and female participants (no interaction).

loading threshold (0.32) for item inclusion First Impressions 13 and interpretation was surpassed only by items 5, (attractive–unattractive), 7 (tolerant–intolerant), and 14 (unmanipulative–manipulative) for Factor B of both the Prediction and Self Indices. A third SI factor was characterized by feelings of self-consciousness, tension, and uncomfortableness with expressing affection. These second PI and SI components did suggest that the attractive versus unattractive dimension might be qualitatively different from the other 13 attributes, but item 5 was nevertheless retained when its removal was found to diminish FIIP test–retest reliabilities in subsequent analyses. JI, PI, and SI scores appear best conceptualized and applied as independent measures generated from the 14-item FIIP rating scale. JI scores were significantly correlated to the JR index, $r (240) = -0.44$, $P < 0.001$, and JVI, $r (240) = 0.46$, $P < 0.001$. The JR and JVI were also inversely related, $r (240) = -0.41$, $P < 0.001$. Each of these three FIIP variables appears to measure unique aspects of first impression formation.

3.6.2. Reliability analyses

JI internal consistency reliabilities of 0.94, 0.93 and 0.94 were found for the 68 men, 184 women, and 252 participants respectively. PI alphas of 0.94, 0.92, and 0.93 were found for the men, women and total group analyses. SI coefficients were 0.89, 0.91 and 0.90 for the same groups respectively.

A certain level of FIIP session to session score changes is unavoidable because of the inevitable partner and interaction differences which will occur from one session to another, but it was also assumed that six-partner FIIP scores would show greater stability as trait measures than those generated from a single session. Weak relationships (correlations ranging from 0.23 to 0.65) between session one and two FIIP measures were indeed found for both men and women, and subsequent analyses indicated that substantial discrepancies were observed among a subset of about 15% of the total sample. Exclusion of participants with JI score differences in excess of 1.1 and 1.2 for men and women, respectively, produced three partner JI, PI, SI, and JR scores that were fairly stable from session to session (average correlations of 0.73, 0.72, 0.68, and 0.54, respectively). First session three-partner FIIP scores were indeed highly correlated $r > 0.80$ in all cases) with the six-partner totals, and test-retest correlation strengths did not differ based upon interval periods which ranged randomly from 1 to 8 weeks in duration.

The temporal stability of six partner (two sessions combined) FIIP scores was difficult to establish since it required participants to complete four separate interaction sessions. Data were collected from a sample of 18 women who completed four sessions with at least five total (and two per session) unfamiliar partners per two sessions (one + two versus three + four). Combined two session (five or six partners) FIIP measures were stable across time with test–retest reliabilities ranging from 0.85 to 0.94. PI and JVI scores were found to decrease similarly for men and women from session one to two in the normative sample, but only small JI score differences were found for this independent sample between sessions two and four, $F (3,15) = 3.14$, $P = 0.03$. Significant differences across the four sessions were not found for predictions, $F (3,15) = 1.70$, $P = 0.18$, judgment variability, $F (3,15) = 2.31$, $P = 0.09$, self ratings, $F (3,15) = 1.83$, $P = 0.15$, and the judgment ratio, $F (3,15) = 1.4$, $P = 0.25$. Fourth session scores did trend toward less critical, less variable, and more confident predictions and self ratings which would warrant consideration if the FIIP was applied repeatedly across time for some other research purpose. The first session partner gender configuration effect on JI scores observed in the larger sample was not replicated, $F (3,14) = 0.25$, $P = 0.86$.

There may be some benefit to classifying participants in future studies into extreme groups on the basis of FIIP judgment scores. High and low classification thresholds can be derived using ± 1S.D from the mean for each judgment variable (Tables 1–3). These classification thresholds essentially assure complete group separation, while median splits would include participants who shift from high to low groups upon retesting. Six-partner classification reliability (identification in same extreme group twice) was 86, 20, and 43% for JI, JR, and JVI scores, respectively. Classification reliability was predictably lower for three partner scores in the normative distribution. Men generated classification reliabilities of 40, 26, and 70% for JI, JR, and JVI scores, respectively, while the percentages of classification agreement among women was 70, 46, and 49% for these same variables. Transpositions were similarly uncommon for classifications based upon single session FIIP scores using 1S.D. (0% for all cases except women classified by the JVI where the rate was 6%).

3.7. Consensus accuracy

The possibility that high FIIP JI scores merely reflected accurate and unbiased appraisals of one or more undesirable partners who were randomly assigned to particular interaction groups warranted consideration. The desirability of individual group partners can be assessed by averaging JI ratings provided by the remaining four group members. High levels of agreement between participant and partner judgments of the same individual might suggest a consensually accurate and unremarkable appraisal of undesirable target characteristics. High levels of FIIP consensus accuracy was not expected given prior impression formation research, and the 14 men and 27 women with elevated (>1 S.D.) JI scores did differ considerably from their partners in the harshness of their social judgments. JI scores provided by the men in this sample were more critical than partner ratings 84% of the time, and male participants with high JI scores generated ratings which averaged 2.15 SDs higher than the mean ratings provided by the remaining group members. JI scores were not significantly correlated with the average ratings of the remaining group members in any of the 14 cases. JI scores provided by the 27 women were also more critical than partner ratings 81% of the time, and female participants with high JI scores generated ratings which averaged 1.66 S.D.s higher than the mean ratings provided by the remaining group members. Participant and partner ratings were significantly correlated in only two of the 27 cases. Elevated JI scores appear to predict ratings that are 1.5 to 2 times harsher than about 80% of other judges indicating consensus inaccuracy.

JR scores differed significantly between the 41 high (>1 S.D.), 46 low (<1 S.D.) and 157 remaining control participants classified by the JI, $F(2,241) = 22.39$, $P = 0.0001$. JR scores among the high group ($M = 0.88$, S.D. $= 0.11$) contrasted sharply with the 46 low ($M = 1.16$, S.D. $= 2.23$) and 157 controls ($M = 1.01$, S.D. $= 0.20$). Conversely, JI group differences were not found in partner judgments of participants, $F(1,241) = 1.34$, $P = 0.26$, suggesting that judgmental participants were not eliciting negative reactions from the partners that they subsequently chose to retaliate against.

3.8. Concurrent validation

A separate sample of 41 Caucasian participants (85% women) selected from the same college population completed at least one FIIP session to provide preliminary evidence of concurrent

validity. Lambert, Cronen, Chasteen, and Lickel (1996) provided a 22-item, seven point Likert scale which was used to measure social judgments of an imaginary 20-year-old black college student. The Lambert procedure was implemented after collection of three-partner FIIP data. Lambert scores were correlated with the three-partner JI, r (39) = 0.35, P = 0.03, JR, r (39) = −0.34, P = 0.03, and JVI, r (39) = 0.38, P = 0.01, measures. Six-partner JI scores were significantly correlated with criticism (averaged over four exit questionnaire items) of the FIIP itself, r (242) = 0.26, P < 0.001.

3.9. Ethnicity considerations

The Native American participants (seven men and seven women) were compared with the remaining white sample in six separate 2 (race) × 2 (gender) ANOVAs to determine if their ethnicity warranted exclusion in the present normatization analyses. Race by gender interactions were not found, and these Native American participants (primarily Ojibwa or Chippewa) did not differ from the white group in JI, F (1,147) = 2.59, P = 0.11, PI, F (1,147) = 1.09, P = 0.30, SI, F (1,147) = 0.83, P = 0.36, JR, F (1,147) = 1.65, P = 0.20, or JVI, F (1,147) = 3.98, P = 0.05, scores. Native American participants did not receive harsher, or more lenient, judgments from their white partners, F (1,147) = 0.03, P = 0.85.

4. Discussion

One objective of the present study was to demonstrate that reliable first impression measures can be generated from a largely uncontrolled and quasi-naturalistic interaction procedure. Men did provide harsher ratings (10/14 items, about 0.5 S.D.s) of both self and others, while predictions became more favorable and judgments less variable over repeated sessions for both male and female participants. Gender differences were not observed for the judgment ratio or variability index, and six-partner FIIP scores did not appear to vary significantly as a function of the gender balance of interaction groups. Balanced (two men and two women) first session groups tended to produce harsher judgments from both men and women, but this configuration effect was not replicated in the second session, in a subsequent independent sample, or in the six-partner analyses. Random group assignment and combined data from multiple sessions appeared to provide internally consistent, temporally stable scores within the desired quasi-naturalistic context without significant contamination by uncontrolled sources of gender variance.

Some researchers may gain extra assurance regarding the test–retest reliability of six-partner FIIP scores by eliminating the roughly 15% of participants who show large inter-session JI differences (1.1 points for men and 1.2 for women). While this control will significantly increase FIIP first and second session correlations, reliability analyses of six-partner scores suggests that this additional control is not necessary. Single session and six-partner FIIP scores were found to be highly correlated. Categorical groups defined by the top and bottom of each judgment variable might also warrant consideration as future independent variables. A 1 S.D. threshold will generate reasonable classification reliability as defined by modest rates of category agreement but minimal risk of group transpositions if participants are retested. Six-partner FIIP scores appear to provide fairly stable trait measures of social judgment, while interpretive caution of individual

session results is warranted by their weaker temporal stability. Individual session scores might be useful to identify extreme participants, conduct exploratory preliminary analyses of certain effects, or to advance understanding of how these FIIP social judgment variables are changed by repeated testing over time.

The group of 41 participants distinguished by tendencies toward routinely harsh (> 1 S.D.) first impressions was particularly interesting. JI ratings by these judgmental participants were much harsher than those rendered by other observers in their respective groups. JR scores were also significantly lower ($P < 0.0001$) among these participants, indicating that their self image and predictions were favorable in comparison with the impressions formed of their six partners. Those partners who received harsh judgments from the group were unremarkable in their own rating tendencies of others. In today's society we are routinely afforded the opportunity to rate and evaluate others with only the benefit of superficial information and brief personal encounters to validate the impression formation process. The present study suggested that most people approach such rating tasks judiciously and select an anchor point just on the favorable side of the center to avoid undue expressions of criticism or support that may later prove to be unwarranted. Some individuals, however, are clearly prepared to forge ahead with judgments that seem to be compelled, as Cronbach might have said, from within rather than without. In this regard, the FIIP was able to identify evidence of a judgmental personality trait. Self-image as described in social identity theory may have provided an important standard in these evaluations since group memberships and categorizations would not have been easily derived or inferred from these relatively brief FIIP encounters.

Social judgment theory warned that ethnic identification might influence the ratings of Native American participants toward their partners or prompt harsher judgments from participants assigned to their interaction groups. The small sample of 14 Native American participants in this study did not, however, differ significantly from the remaining Caucasians in their social judgments, self-perceptions, or the impressions they left on their partners. The trend for less variable Native American judgments of their white partners also differed from expectations (Doosje et al., 1998) of higher variability in outgroup (minority) judgments of ingroup (majority) partners. It was the similarities, rather than differences, that were noteworthy.

Finally, exit questionnaire appraisals of the FIIP were decidedly favorable, and negative comments were infrequent ($< 1\%$). College students seem to enjoy the opportunity to meet and interact. Exclusion rates can be reduced substantially by merely asking participants to identify familiar classmates in a space on the consent form prior to group assignments.

Appendix A

FIIP rating form[1]

Examine the 14 general characteristics or personality features identified below. Please use this rating form to, first, describe each of your interaction partners. Secondly, try to predict how each of your partners rated you. Finally, how would you rate yourself in terms of each item. Pleaseuse the rating scale provided in the laminated booklet. Items left blank will be scored "4" to indicate the absence of a strong impression.

	YOUR DESCRIPTION OF EACH OF YOUR THREE PARTNERS	YOUR PREDICTION OF EACH OF YOUR THREE PARTNERS	PERSONAL SELF IMAGE
ITEM	A B C D	A B C D	
1. good vs poor listener	—	—	—
2. good vs poor humor	—	—	—
3. interesting vs boring	—	—	—
4. intelligent vs unintelligent	—	—	—
5. attractive vs unattractive	—	—	—
6. optimistic vs pessimistic	—	—	—
7. tolerant vs intolerant	—	—	—
8. affectionate vs unaffectionate	—	—	—
9. perceptive vs unperceptive	—	—	—
10. flexible vs rigid	—	—	—
11. authentic vs inauthentic	—	—	—
12. relaxed vs tense	—	—	—
13. assured vs self-conscious	—	—	—
14. unmanipulative vs manipulative	—	—	—

---

[1] FIIP test kits (reusable color-coded booklets and forms) and BASIC scoring program is available from: Alan R. King, Ph.D., Department of Psychology, University of North Dakota, PO Box 8380, Grand Forks, North Dakota, 58202-8380, USA (or alan_king@und.nodak.edu).

Participants provided a laminated copy of the items anchored on the seven point rating scale. Column assigned to participant (A, B, C or D) colored out on form prior to distribution.

References

Aronson, E., & Linder, D. E. (1965). Gain and loss of esteem as determinants of interpersonal attractiveness. *Journal of Experimental Social Psychology*, *1*, 156–171.

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

Asch, S. E. (1952). *Social psychology*. Englewood Cliffs, NJ: Prentice-Hall Publishers.

Baum, A., Fisher, J. D., & Singer, J. E. (1985). *Social psychology*. N.Y: Random.

Berry, D. S., Hansen, J. S., Landry-Pester, J. C., & Meier, J. A. (1994). Vocal determinants or first impression of young children. *Journal of Nonverbal Behavior*, *18*(3), 187–197.

Birdwhistell, R. L. (1970). *Kinesics and context: essays on body motion communication*. Philadelphia: University of Pennsylvania Press.

Bryan, L. R., Coleman, M., Ganong, L. H., & Bryan, S. H. (1986). Person perception: Family structure as a cue for stereotyping. *Journal of Marriage and the Family*, *48*, 169–174.

Cash, T. F., Gillen, B., & Burns, D. S. (1977). Sexism and ''beautyism'' in personnel consultant decision making. *Journal of Applied Psychology*, *62*, 301–310.

Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The Interpersonal Perception Task. *Journal of Nonverbal Behavior*, *13*(4), 225–245.

Cronbach, L. J. (1955). Processes affecting scores on ''understanding of others'' and assumed similarity. *Psychological Bulletin*, *52*(3), 177–193.

DiGirolamo, G. J., & Hintzman, D. L. (1997). First impressions are lasting impressions: a primacy effect in memory for repetitions. *Psychonomic Bulletin and Review*, *4*(1), 121–124.

Dittmar, H. (1992). Perceived material wealth and first impressions. *British Journal of Social Psychology*, *31*(4), 379–391.

Doise, W., Deschamps, J. C., & Meyer, G. (1978). The accentuation of intra-category similarities. In H. Tajfel (Ed.), *Differentiation between social groups*. London: Academic Press.

Doosje, B., Haslam, S. A., Spears, R., Oakes, P. J., & Koomen, W. (1998). *European Journal of Social Psychology*, *28*, 173–184.

Doosje, B., Spears, R., & Koomen, W. (1995). When bad isn't all bad: the strategic use of sample information in generalization and stereotyping. *Journal of Personality and Social Psychology*, *69*, 642–655.

Dougherty, T. W., Turban, D. B., & Callender, J. C. (1994). Confirming first impressions in the employment interview: a field study of interview behavior. *Journal of Applied Psychology*, *79*(5), 659–665.

Eiser, J. R., & Stroebe, W. (1972). *Categorization and social judgement*. London: Academic Press.

Frey, K. P., & Smith, E. R. (1993). Beyond the actor's traits: Forming impressions of actors, targets, and relationships from social behaviors. *Journal of Personality and Social Psychology*, *65*(3), 486–493.

Hart, A. J., & Morry, M. M. (1997). Trait inferences based on racial and behavioral cues. *Basic and Applied Social Psychology*, *19*(1), 33–48.

Jones, E. E., Rock, L., Shaver, K. G., Goethals, G. R., & Ward, L. M. (1968). Pattern of performance and ability attribution: an unexpected primacy effect. *Journal of Personality and Social Psychology*, *10*, 317–340.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurements*, *20*, 141–151.

Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: a social relations analysis. *Psychological Bulletin*, *102*(3), 390–402.

Kenny, D. A., Albright, L., Malloy, T. E., & Kashy, D. A. (1994). Consensus in interpersonal perception: acquaintance and the big five. *Psychological Bulletin*, *116*(2), 245–258.

Kenny, D. A., & DePaulo, B. M. (1993). Do people know how others view them? An empirical and theoretical account. *Psychological Bulletin*, *114*(1), 145–161.

Lambert, A. J., Cronen, S., Chasteen, A. L., & Lickel, B. (1996). Private vs public expressions of racial prejudice. *Journal of Experimental Social Psychology*, *32*, 437–459.

Levesque, M., & Kenny, D. A. (1993). Accuracy of behavioral predictions at zero acquaintance: a social relations analysis. *Journal of Personality and Social Psychology*, *65*(6), 1178–1187.

Lydon, J. E., Jamieson, D. W., & Zanna, M. P. (1988). Interpersonal similarity and the social and intellectual dimensions or first impressions. *Social Cognition*, *6*(4), 269–286.

Mette, D. R. (1971). Rejection of unexpected success as a function of the negative consequences of accepting success. *Journal of Personality and Social Psychology*, *17*, 332–341.

Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: an integration. *European Journal of Social Psychology*, *22*, 103–122.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.

Rosenberg, S., & Olshan, K. (1970). Evaluative and descriptive aspects in personality perception. *Journal of Personality and Social Psychology*, *16*, 619–626.

Sherif, M., & Hovland, C. (1961). *Social judgement: assimilation and contrast effects in accomodation and attitude change*. New Haven: Yale University Press.

Spears, R., & Manstead, A. (1989). The social context of stereotyping and differentiation. *European Journal of Social Psychology*, *19*, 101–121.

Steele, K. M., & Smithwick, L. E. (1989). First names and first impressions: A fragile relationship. *Sex Roles*, *21*(7-8), 517–523.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd Ed.). New York: Lawrence Erlbaum Associates, Publisher.

Taifel, H. (1978). *Differentiation between social groups*. London: Academic Press.

Taifel, H., & Turner, J. (1986). The social identity theory of intergroup conflict. In S. Worchel, & W. Austin (Eds.), *Psychology of intergroup relations*. Chicago: Nelson-Hall.

Taifel, H., & Wilkes (1963). Classification and quantitiative judgement. *British Journal of Social Psychology*, *54*, 101–114.

Temple, L. E., & Loewen, K. R. (1993). Perceptions of power: First impressions of a woman wearing a jacket. *Perceptual and Motor Skills*, *76*(1), 339–348.

Turner, J. C., Hogg, M., Oakes, P. J., Reicher, S., & Wetherall, M. (1987). *Rediscovering the social group: a self-categorization theory*. Oxford: Basil Blackwell.

Walther, J. B. (1993). Construction and validation of a quantitative measure of impression development. *The Southern Communication Journal*, *59*(1), 27–33.