A Model of Language-Based Impression Formation and Attribution Among Germans

Tobias Schröder

The online version of this article can be found at:
http://jls.sagepub.com/content/30/1/82

Published by:
$SAGE

http://www.sagepublications.com

Additional services and information for *Journal of Language and Social Psychology* can be found at:

Email Alerts: http://jls.sagepub.com/cgi/alerts

Subscriptions: http://jls.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://jls.sagepub.com/content/30/1/82.refs.html

A Model of Language-Based Impression Formation and Attribution Among Germans

Tobias Schröder[1]

Abstract

Research on the English and the Japanese languages has shown that impression formation and attribution processes can be modeled as stemming from a desire to maximize affective coherence in linguistic representations of social events. This article describes a replication in the German language, simultaneously uncovering subtle cultural distinctions. Subjects ($N = 1,905$) rated 376 nouns, 393 verbs, 331 adjectives, 128 combinations of nouns and adjectives, and 100 short descriptions of social interaction on the three dimensions of the semantic differential (evaluation, potency, and activity). The data were used to estimate a set of regression equations that can be used to model impression formation and attribution. Sample applications of the model demonstrate its ability to predict the outcome of textbook classics in experimental social psychology.

Keywords

impression formation, attribution, semantic differential, language, affect control theory

Imagine how a mother plays with her daughter. Great, you may think (or unthinkingly react). That is one of the things that mothers should be doing with daughters. Unfortunately, life is not always that much in order, so you might stumble on a mother who has just beat her daughter. Would such an event not make you feel uncomfortable and goad your mind for an explanation? You might infer the mother to be in a furious

[1]Freie Universität Berlin, Berlin, Germany

Corresponding Author:
Tobias Schröder, University of Waterloo, Department of Philosophy (HH143), 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1
Email: mail@tschroeder.eu

emotional state or to be a violent person. On the other hand, what can you expect a daughter beaten by her mother to be like? Submissive or uncertain might be good guesses about her personality. To see a furious mother beat a submissive daughter is certainly still not an enjoyable experience, but one that may pass your mind without provoking too many questions about the situation.

These introductory predictions about your emotional reactions and cognitive inferences based on short descriptions of social actions were derived from a mathematical model of the affective processing of language that the present article deals with. The model was developed with data from German-speaking subjects, adding empirical evidence to existing research in American English (Averett & Heise, 1987; Smith-Lovin, 1987) and Japanese (Smith, Matsuno, & Ike, 2001; Smith, Matsuno, & Umino, 1994), which has shown that the affective structure of the language shaping people's social cognitions may account for a variety of social-psychological phenomena (for review, see Heise, 2007).

Impression Formation and Attribution

Observing other persons' behavior is an important source of information for understanding them. According to correspondent inference theory (Jones & Davis, 1965), people try to infer from an action whether the act itself corresponds to an enduring personality characteristic of the actor. Along with perceptions of the actor's freedom of choice and the intended outcomes of the behavior in question, it is the expectedness of that behavior that is a key variable for inferring dispositions from acts. In a classic experiment, participants listened to a supposed "job interview" between a psychologist and an applicant for a job that required an extravert versus an introvert personality (Jones, Davis, & Gergen, 1961). The "applicant" acted either consistently or inconsistently with the role-specific personality descriptions. After listening, subjects were asked to rate their impressions of the applicant's "real" personality. They tended to perceive the answers that violated role expectations for the job as more revealing. Remember the introductory example to the present article. One would expect a mother to play with a daughter; hence, no further inference about the mother's personality traits is necessary. In contrast, a mother beating a daughter clearly violates social expectations. Assuming a violent personality trait of the mother provides a possible explanation for this violation.

People infer not only personality traits of the actor from a certain action but also characteristics of the target to which the action is directed. The tendency to disparage victims of misfortune is a widely known social-psychological phenomenon. It is often assumed that poor people are lazy and that crime victims are careless. Moreover, there seem to be social expectations for the vulnerability to crime for certain roles. Attributions of accountability depend on those expectations. In an experiment by Jones and Aronson (1973), victims of rape were faulted the more for the crime the higher their initial social status. If a victim was presented as "virgin" or "married woman" (high status) she was disparaged more than if she was believed to be a "divorcee" (low status).

Apparently, participants expected the action of rape to be more consistent with a "divorcee" as the target as compared with a "married woman." The expectation of specific actions is linked not only to the social role of the actor but also to that of the target. In the introductory example to this article, a daughter was expected to be played with rather than beaten by her mother.

Language-Based Models of Impression Formation and Attribution

Affect control theory (Heise, 1979, 2007; MacKinnon, 1994; Smith-Lovin & Heise, 1988) proposes that people rely on the affective structure of the language to determine whether social expectations are met in specific actions, and, if not, to infer explanatory characteristics of actors and targets. Osgood, Suci, and Tannenbaum (1957) showed that the three dimensions of evaluation (E; good vs. bad), potency (P; strong vs. weak), and activity (A; lively vs. calm) constitute an affective space as a structuring principle of the language. They developed the semantic differential, an instrument to measure the emotional meaning of a given word within the evaluation–potency–activity (EPA) system, based on antonym scales with contrasting adjectives at each end. In an extensive cross-cultural research program, Osgood, May, and Miron (1975) gathered impressive empirical evidence for the contention that the three dimensional structure of the semantic differential is universal and can be found in virtually any natural language. Competing models from diverse domains such as personality (McCrae & Costa, 1989), emotion (Fontaine, Scherer, Roesch, & Elssworth, 2007), or interpersonal relations (Wish, Deutsch, & Kaplan, 1976) may imply differing core dimensions of social experience, but they are all, to some extent, transformable to the EPA structure (Scholl, 2009). For example, extraversion can be thought of as a linear combination of slightly positive evaluation, high potency, and moderate activity (Mehrabian, 1995). Mathematically, it is irrelevant which combination of vectors one chooses as basis vectors to describe a given Euclidean space. Therefore, one might agree on that the semantic differential is a viable approach for describing the ubiquitous dimensional structure of socioemotional perception.

In the affect control theory research program, the semantic differential has been used to compile large affective dictionaries, that is, databases of the most frequently used words in a language along with the average EPA ratings of these words by a sample of native speakers (see Heise, 2001, for a summary of studies in various languages). Whereas psychologists use such dictionaries for scaling the emotional impact of stimuli used in experiments, sociologists have linked them to culture and social structure. For example, Schneider (2004) demonstrated that concepts designating authority (e.g., father, judge, or president) are systematically rated higher on evaluation by Americans than by Germans. His conclusion was that a cultural difference exists between the United States and Germany in that Americans generally have more positive attitudes toward power and authority. This result can be linked to Hofstede's (1984) research in intercultural organizational psychology. Hofstede showed that cultures differ in the amount of power

distance, the extent to which power differences are accepted in a given society. In American society, such power differences are larger than in Germany, as demonstrated by Hofstede's (1984) power distance index. Schneider's (2004) work shows that information about the cultural trait of power distance seems to be contained in the emotional associations of relevant words. Thus, an important implication from the work on affective dictionaries is that affective connotations seem to contain culture-specific information about social expectations bound to certain concepts.

According to affect control theory, people use the affective meaning of words as a source of information to infer whether a given social event corresponds to cultural norms. When social actions are described in ordinary language, these descriptions may differ in the amount of affective coherence, which can be understood as the mutual goodness of fit of all the connotations involved. Affective coherence can be modeled mathematically using impression formation equations, obtained in empirical studies by regressing EPA ratings of words in the context of a sample of given events on out-of-context semantic differential ratings of the same words, available in the affective dictionary (Smith-Lovin, 1987). The squared Euclidean distance between the affective outcome of a social event and the initial sentiments toward the actor, behavior, and object (affective deflection) can be used as a measure of the violation of social expectations (Heise, 2007). "A mother plays with a daughter" is a coherent and, thus, socially expected event because the feeling toward the mother that results from this event is very similar to the fundamental feeling toward any mother out of the context of the event. However, the affective impression toward a mother who has just beat her daughter would be quite different, resulting in a high affective deflection and a desire for an explanation.

Whereas impression formation equations model the affective consequences of social events, a different type of regression equations model the amalgamation of affect resulting from the combination of concepts. The emotional meaning of a modified social identity such as a "violent mother" can be predicted by taking a nonlinear combination of the fundamental EPA ratings of the concepts "violent" and "mother." The regression weights of those equations are derived in empirical studies (Averett & Heise, 1987; Smith et al., 2001). Derivations of the amalgamation equations can be used to calculate the optimal EPA vector of a modifying adjective that must be inserted into a given event description to maximize its affective coherence. "A violent mother beats a daughter" is a coherent event because the affective actor impression resulting from it is very similar to the amalgamated feeling toward a "violent mother" as such.

The Present Study

There is considerable evidence for the ubiquity of the EPA structure of affective connotations; however, it is an open empirical question whether impression formation and attribution can be modeled in similar ways across cultures. Most of the existing research in affect control theory has been conducted with English language material. Another complete model has been estimated for Japanese (Smith et al., 1994; Smith

et al., 2001). The research described here aimed at adding a German model, extending previous research by Schneider (1989) who has gathered semantic differential ratings of roughly 800 German words. An additional goal of the present article is to demonstrate how the model can be applied to predict the outcome of well-known classic social-psychological experiments that deal with impression formation and attribution.

Method

Construction of Stimuli

The model consists of an affective dictionary, nine regression equations modeling impression formation, and three regression equations modeling amalgamation of affect. Hence, three types of stimuli (words, event descriptions, and word combinations) had to be compiled in order to be rated with the Semantic Differential by a German speaking sample.

The *affective dictionary* contains a total of 1,100 words. Of these, 376 nouns designate social identities, 393 verbs represent social behaviors, and 331 adjectives denote modifying characteristics such as personality traits and emotional states. The selection of words for the study aimed at including the most frequently used German words in descriptions of social interaction. A corpus of German words, published on the Internet (Universität Leipzig, 2007), served as an important resource as well as translations of MacKinnon's (1994) compilation of social roles, and adjective-based personality scales (Jacobs & Scholl, 2005) and emotion concepts (Morgan & Heise, 1988; Scherer, 2005). During the rating task, the words appeared in a grammatical form that implied their potential use in descriptions of social events: All the nouns were presented with the indefinite article (e.g., "eine Mutter" [a mother]), whereas the verbs appeared with the indefinite pronoun (e.g., "jemanden loben" [praise someone]), and the adjectives with an auxiliary verb (e.g., "wütend sein" [to be angry]).

To estimate the impression formation equations, the subjects' semantic differential ratings of actor, behavior, and object in the context of sample event descriptions (e.g., "A mother beats a daughter.") were assessed. To ensure that the model would reflect actual impression formation processes rather than an arbitrary semantic configuration of sample event descriptions, a balanced design was required. All the possible combinations of positive and negative values on the evaluation, potency, and activity scales yield eight different configurations of affective word meaning. Thus, a full factorial design would have required $8^3 = 512$ different types of actor–behavior–object sentences (Smith-Lovin, 1987). Such a research design is extremely resource consuming, given that for each event description, nine ratings (actor, behavior, object × E, P, A) from at least 25 to 30 subjects are required (Smith-Lovin, 1987). Therefore, a more economical type of design with only 100 sentences was chosen that had proved its methodological robustness beforehand in a similar study in Japanese (Smith et al., 1994).

Between cultures, the sentiments attached to certain concepts may differ considerably, even though the denotative meaning of the concept is identical in the two

languages involved (Heise, 2001). Therefore, in replicating the Japanese design, a match of the affective configuration of the sentences was sought rather than a direct translation. For example, in the event "A rival plays with the rascal" (Smith et al., 1994, p. 137, Nr. 1) a rival as well as the action of playing are good (E+), strong (P+), and lively (A+), according to the Japanese affective dictionary, whereas a rascal is perceived as negative (E−), strong (P+), and lively (A+). To create a corresponding German event description, words with similar affective configurations were chosen from Schneider's (1989) affective dictionary, resulting in the sentence "Ein Athlet (E+ P+ A+) rettet (E+ P+ A+) einen Rüpel (E− P+ A+)" [An athlete saves a lout]. Each event was presented to the subjects along with a question designating the part of the sentence that ought to be rated (e.g., "An athlete saves a lout. How do you feel about the lout in this situation?"). A list of all the sentences used in the study is available on request from the author.

For the stimuli that elicited the in-context ratings for estimating amalgamation equations, a full factorial design was used. Eight possible types of concepts designating social identities combined with eight types of modifying adjectives yielded 64 configurations of concept combinations. There were two stimuli of each type (hence, 128 in total), as one of the adjectives always referred to an emotional state, whereas another designated a personality trait. Among the emotion–identity combinations was "eine verärgerte (E− P+ A+) Dame (E+ P− A−)" [an angry lady]; an example for the trait–identity combinations is "ein bescheidener (E+ P− A−) Fischer (E+ P+ A−)" [a modest fisherman]. A list of all the stimuli is again available on request from the author.

Semantic Differential Scales

The ratings were obtained via the Internet with SURVEYOR (Heise, 2001), a Java applet that allows the subjects to rate the presented concepts by dragging a pointer along bipolar adjective scales with the computer mouse. The contrasting adjectives were the same as in Schneider's (1989) study (*for E:* angenehm [pleasant], gut [good], schön [beautiful], freundlich [friendly] *versus* unangenehm [unpleasant], schlecht [bad], hässlich [ugly], unfreundlich [unfriendly]; *for P:* groß [large], schwer [heavy], stark [strong], kraftvoll [powerful] *versus* klein [small], leicht [light], schwach [weak], zart [gentle]; *for A:* schnell [quick], geräuschvoll [noisy], bewegt [eventful], lebhaft [lively] *versus* langsam [slow], still [quiet], ruhig [calm], träge [inert]). Skrandies (1998) provides evidence for the orthogonal EPA structure of these scales. The middle position on the scale was labeled neutral, and reference positions going outward were labeled with adverbs such as slightly, quite, very, and extremely. Depending on the position of the pointer on the scale, the ratings were coded with continuous scores ranging from −4.3 for extremely bad, weak, or calm through 0 for neutral to +4.3 for extremely good, strong, or lively. For economy, only one rating per stimulus was obtained for each of the dimensions.

Procedure and Sample

There were 1,905 informants (734 male and 1,171 female) from all over Germany who took part in the Internet-based study. The research had been advertised as a "study of language and emotion" in an extensive campaign, including mailing lists from different universities, weblogs, newspaper reports, and radio interviews. Most of the participants ($N = 1,029$) were between 20 and 29 years of age, but the sample covered all ages, including $N = 129$ younger than 20 years and $N = 92$ older than 60 years. Data from 83 persons (4.4%) were excluded from the analysis, as they had indicated that German was not their mother tongue. Interest in the topic of language and emotion can be considered the subjects' main motivation for taking part in the study. There was a 200 euro (about US$300) lottery as an appreciation of their efforts. The data collection took place between February and October, 2007.

Once a participant had gone through the instructions on the web page and clicked on the button to start the rating task, the SURVEYOR applet randomly chose one out of 24 files that each contained 40 to 42 single concepts, 12 to 13 sentences, and 5 to 6 concept combinations. Each participant thus rated about 60 stimuli. The median duration of the task was 23 minutes. Because of the random procedure, the exact number of raters per stimulus ranged from 17 male and 22 female raters to 46 male and 54 female raters. On average, each stimulus was rated by 29.5 male and 46.4 female raters.

Results

Cultural Consensus

The research aimed at assessing the cultural consensus rather than individual differences among Germans. Therefore, principal component analyses based on *q* correlations were conducted with each of the 24 data sets. *Q* correlations indicate to what extent two *persons* agree on their ratings of all the items as opposed to *r* correlations that are more common in psychological research and that refer to the similarity of *items*. A similar pattern of extracted factors emerged in all the data sets. One huge factor could be extracted, usually accounting for 50% to 60% of the variance in ratings. This factor reflects what all the subjects had in common when doing the ratings; hence, it can be interpreted as a cultural base underlying the individuals' judgments (see Romney, Weller, & Batchelder, 1986, on culture-as-consensus theory).

In psychometric theory, items with a low item-total correlation are usually excluded from the test to enhance reliability. Similarly, subjects whose EPA ratings barely reflect the common cultural factor may be discarded from the analysis of culturally shared affective word meaning (Heise, 2010) to get more culturally adequate results. The commonality h^2 can be used as a measure of the cultural accuracy of a person's ratings. Finding a relevant criterion for discarding a participant's data from the analysis required considerable trial and error, but eventually, the reliability of the ratings

seemed to be optimal when only considering subjects with a commonality of at least $h^2 = 0.30$. Accordingly, the data from 202 subjects were discarded, finally resulting in an average of 25.6 male and 37.8 female raters per stimulus. The affective dictionary that is based on the average evaluation, potency, and activity ratings of those remaining raters is available on request from the author.

Reliability of the Ratings

Estimations of the reliability of the ratings were derived in an analogy to the concept of split half reliability from psychometric theory (see Lienert & Raatz, 1994). For each of the 24 data sets, the ratings from randomized subsamples were correlated. To compensate for the smaller sizes of the subsamples, the correlations were corrected using the Spearman–Brown formula that is employed in psychometrics. The resulting reliability estimates ranged from $r = .95$ to $r = .99$ for evaluation, from $r = .76$ to $r = .98$ for potency, and from $r = .88$ to $r = .99$ for activity. For the male raters, the median reliability estimates were $r = .98$ (E), $r = .94$ (P), and $r = .96$ (A). The corresponding median reliability estimates for the female raters were $r = .99$ (E), $r = .97$ (P), and $r = .97$ (A). The correlations between male and female ratings were $r = .98$ (E), $r = .92$ (P), and $r = .93$ (A). These are only slightly below the reliability estimates, what points to a large consensus between men and women about the affective meaning of words.

Estimating Impression Formation Equations

There are nine impression formation equations: a separate one for evaluation, potency, and activity for actor, behavior, and object. They have the following general form:

$$A' = \text{Constant} + \beta_1 A + \beta_2 B + \beta_3 O + \text{Interaction terms},$$

where A' is the predicted in-context impression of the actor, and A, B, and O are the out-of-context EPA ratings of actor, behavior, and object, respectively, taken from the affective dictionary. Additional interaction terms account for psychological subtleties in the impression formation process. For example, the term $B_e O_e$ refers to a principle of evaluative balance (nice objects should be treated nicely, whereas bad objects can be treated poorly), which can be derived from Heider's (1946) balance theory. Note that in the equations, subscript letters e, p, and a denote Osgood's affective dimensions: evaluation, potency, and activity. For example, O_p corresponds to the potency impression of the object.

The regression weights were estimated using an ordinary least squares routine. The subject's average in-context EPA ratings of actor, behavior, and object from the 100 sample event descriptions served as criteria, whereas the corresponding EPA ratings from the affective dictionary were used as predictors. Additional predictors were 27 two-way interaction effects incorporated into the regression as product terms (e.g., $B_e O_e$) and 27 three-way interaction effects (e.g., $A_e B_p O_a$).[1]

The procedure required attention to the issue of multicollinearity, as the product terms were expected to substantially inflate the variance of their components. Therefore, a stepwise routine was chosen for estimating the equations. Step by step, only those additional predictors were included into the model (a) that were statistically significant ($p < .05$), (b) whose tolerance value was $1 - r^2 > .50$, and (c) that significantly added to the predictive power of the overall model ($p < .01$). Tentative estimations of the equations with different routines (SPSS enter and backward procedures) yielded very similar results but seemed to be affected largely by multicollinearity problems. To test for the impact of the semantic configuration of the sample events on the structure of the equations, the estimation procedure was repeated three times with random subsamples of 50 events, resulting in very similar models whose predicted values had high correlations with the predicted values of the original estimations ($.88 < r < .98$; median $r = .95$). As a consequence, the equations described below can be considered quite robust and mainly reflecting psychological processes rather than methodological artifacts. Table 1 lists all the coefficients. The equations were all statistically significant ($p < .001$) and accounted for between 61.5% and 83.3% of the variance.

Among the most important predictors are those that can be interpreted as stability effects. Good actors retain much of their positive pre-event evaluation, regardless of what happens during the event. Potent behaviors retain a lot of their original strength, no matter who inflicts them on whom, and so forth. These coefficients can be linked to the conservative tendencies of the human mind. To a certain extent, people stick to their fundamental feelings about others even when they perceive contradicting actions. But of course, social behavior does matter in impression formation. The equations also point to actors being seen in light of their actions (and vice versa). Good behaviors make both actors and object persons seem good. For the prediction of post-event actor potency impressions, the behavior effect even seems to excel stability ($.57 * B_p$ as opposed to $.39 * A_p$). People perceive those as strong who behave dominantly. Correspondingly, object persons exposed to potent behaviors seem rather weak in light of the event ($-.54 * B_p$). The power differential between actors and objects is of mutual importance for the potency impression: The stronger the actor, the weaker the in-context impression of the object, and vice versa.

Several interaction terms can be interpreted in terms of psychological consistency. As Heider's (1946) balance theory would predict, actors are seen in a better light if their actions are consistent with the fundamental evaluative feelings associated with themselves ($.05 * A_e B_e$) and with the object ($.04 * B_e O_e$). That is, the effect of the behavior on the actor impression is moderated by situational expectations. The same is true for the activity dimension. Performing lively behaviors makes actors seem lively ($.52 * B_a$), but less so, if the actors were already perceived as lively before the act ($-.06 * A_a B_a$). There are also some important cross-dimensional interaction effects. Acting powerfully on a good person (e.g., beating a daughter) makes an actor seem much less nice ($-.13 * B_p O_e$). Playing a major role in five of the equations, this interaction between behavior potency and object evaluation seems to be an especially important psychological process in impression formation among Germans.

Table 1. Equations for Predicting Impressions Resulting From Event Descriptions

| | Coefficients for Predicting In-Context | | | | | | | | |
| | Impressions of Actor | | | Impressions of Behavior | | | Impressions of Object | | |
Term	Evaluation	Potency	Activation	Evaluation	Potency	Activation	Evaluation	Potency	Activation
Constant	−.38	−.03	.10	−.72	−.05	.18	−.15	−.26	−.57
A_e	.42			.23					
A_p		.39			.17		.10	−.28	−.18
A_a	−.11	.08	.39		.10	.28			
B_e	.47	−.07	−.13	.51		−.06	.13	.17	
B_p		.57	.14		.66			−.54	
B_a			.52			.62		.15	
O_e	.11			.20			.38		
O_p		−.20						.40	
O_a		.16							.28
$A_e B_e$	.05			.06		−.02	.06		
$A_e B_p$				.08					
$A_e B_a$					.02				
$A_e O_e$				.04			.03		
$A_e O_p$				−.04				.03	
$A_e O_a$	.06			.04					
$A_p B_p$							−.04		
$A_p B_a$		−.04	−.03					.08	.05
$A_p O_e$			−.03			−.03			
$A_p O_p$									.05
$A_a B_e$							−.03		.03
$A_a B_p$					−.09				
$A_a B_a$			−.06			−.07			
$A_a O_e$	.09							.09	
$A_a O_p$		−.07	.04	.05			.04	.06	
$A_a O_a$	.09			.09					
$B_e O_e$	.04			.06					
$B_e O_p$						.04			
$B_e O_a$	−.07			−.09	−.05	.04			
$B_p O_e$	−.13			−.10	.02			−.06	−.08
$B_p O_p$			.07						.08
$B_p O_a$							−.06		
$B_a O_e$		.03							
$B_a O_p$		.06				.08			
$A_e B_e O_p$	−.03								
$A_e B_e O_a$									.01
$A_e B_p O_e$	.02								
$A_e B_p O_a$		.02		.03					
$A_e B_a O_e$						.02			.01
$A_e B_a O_p$					−.01				
$A_p B_e O_a$	.03								

(continued)

Table 1. (continued)

| | Coefficients for Predicting In-Context | | | | | | | | |
| | Impressions of Actor | | | Impressions of Behavior | | | Impressions of Object | | |
Term	Evaluation	Potency	Activation	Evaluation	Potency	Activation	Evaluation	Potency	Activation
$A_p B_p O_e$	−.02								
$A_p B_p O_p$									−.03
$A_p B_p O_a$					.02				
$A_p B_a O_e$						.02			
$A_p B_a O_a$		.02						−.03	
$A_a B_e O_a$					.03				−.03
$A_a B_a O_p$			−.04	−.05		−.03			
$A_a B_a O_a$						.03			−.02
R^2	.830	.779	.816	.830	.751	.833	.776	.747	.615

Note: A = actor; B = behavior; O = object; e = evaluation; p = potency; a = activity. Only significant predictors are presented (at least $p < .05$).

Estimating Amalgamation Equations

The procedure for estimating the amalgamation equations was essentially the same as for the impression formation model. The evaluation, potency, and activity ratings of identity–adjective combinations were regressed on the EPA ratings of the components. The equations have the following general form:

$$C' = \text{Constant} + \beta_1 I + \beta_2 M + \text{Interaction terms}$$

where C' is the predicted feeling elicited by the concept combination, whereas I and M are the EPA ratings of the corresponding words from the affective dictionary, denoting identities and modifying personality traits or emotional states, respectively. A stepwise ordinary least squares routine was used to determine the coefficients, allowing for constants, all six main effects and nine two-way interaction effects.[2] The equations were all statistically significant ($p < .001$) and accounted for between 78.1% and 86.4% of the variance. The coefficients of the resulting amalgamation model are listed in Table 2.

For all three dimensions of the semantic differential, the affective connotations of the modifying adjective seem to have a more important effect on the amalgamated affect than the emotional meaning of the noun. Apparently, individualized expectations based on perceived personality traits and emotional states have greater impact on impression formation than the mere social roles. For example, the perception that a mother is "angry" weighs more in judgments of her behavior than the fact that she is a "mother." There are also some meaningful cross-dimensional main effects (e.g., addition of a potent modifier reduces the evaluation of the concept combination) and

Table 2. Equations Predicting Amalgamated Feelings Toward Concept Combinations

Term	Coefficients for Predicting Amalgamated		
	Evaluation	Potency	Activation
Constant	−.50	−.36	.09
I_e	.39	.09	
I_p	.06	.48	
I_a		.07	.36
M_e	.60		−.12
M_p	−.10	.60	.11
M_a	−.13		.59
$I_e M_e$	.08		
$I_e M_a$	.02		
$I_p M_a$	−.05		−.05
$I_a M_a$			.07
R^2	.864	.781	.838

Note: I = identity concept; M = modifying adjective; e = evaluation; p = potency; a = activity. Only significant predictors are presented (at least $p < .01$).

balance effects (e.g., if a good identity is modified with a good adjective, there is an additional evaluative bonus).

Sample Applications of the Model

The sum of squared differences between out-of-context sentiments toward, and transient in-context affective impressions of, the actor, behavior, and object of a social event can be used as a numerical index of affective deflection in the event description (Heise, 2007):

$$D = (A'_e - A_e)^2 + (A'_p - A_p)^2 + (A'_a - A_a)^2 + (B'_e - B_e)^2 + (B'_p - B_p)^2 + \\ (B'_a - B_a)^2 + (O'_e - O_e)^2 + (O'_p - O_p)^2 + (O'_a - O_a)^2$$

where A, B, and O denote pre-event EPA ratings of actor, behavior, and object of the described social action, respectively; whereas A', B', and O' denote the post-event affective impressions that can be calculated with the impression formations displayed in Table 1.

Affective deflection is inversely related to the expectedness of a social event; hence, descriptions of events corresponding to social norms should have low D values, whereas norm-violating social actions should result in a high D.

Mothers and daughters. In the introductory example to the present article, a mother plays with a daughter. Inserting the average (female) EPA ratings of mother ($A_e = 2.9$,

$A_p = 1.5$, $A_a = 0.6$), to play with ($B_e = 2.8$, $B_p = 0.8$, $B_a = 2.2$), and daughter ($O_e = 2.8$, $O_p = -0.9$, $O_a = 0.9$) from the affective dictionary into the impression formation equations (Table 1) and in the deflection equation yields a low deflection value of $D = 2.8$. In contrast, the event "a mother beats a daughter" was introduced as an example of an affectively incoherent event that stimulates the search for possible explanations. Substituting the verb to play with to beat ($B_e = -3.9$, $B_p = 2.5$, $B_a = 2.9$) in the equations results in a very high deflection value of $D = 43.8$. To facilitate the calculations, the equations have been incorporated into affect control theory's INTERACT software (Heise, 1997). A German version of INTERACT can be used online at the author's personal website.[3]

What kind of mother would beat her daughter? To answer this question using the attribution model described in the present article, the amalgamation equations (Table 2) must substitute the actor-designating terms of the impression formation equations (Table 1) and the deflection equation. The resulting set of equations must be solved for an optimal evaluation–potency–activity vector of the modifying adjective that allows the deflection to be minimal. Fortunately, INTERACT can be used to do the necessary calculations. In the example, the adjective that would minimize the deflection should have average ratings of $E = -4.6$, $P = 4.4$, and $A = 3.2$. No adjective with exactly that profile is contained in the German affective dictionary. The closest ones (in terms of Euclidean distances) are violent and aggressive (trait attributions) or irate and furious (emotion attributions). In the same way, the model can be used to make inferences about the daughter in light of this event, substituting the object-related terms in the impression formation and deflection equations with the amalgamation equations. The resulting suggestions for adjectives to be added to the word "daughter" in order to maximize affective coherence are submissive and uncertain (traits) or afraid and fearful (emotions). It follows that the attribution process during a social event can be modeled plausibly by computationally maximizing the affective coherence of the event description within the three-dimensional affective space.

Correspondent inference theory. In the classic Jones et al. (1961) experiment, participants considered only those behaviors of a supposed job applicant during a job interview as truly revealing that violated introvert versus extravert role expectations. The experiment can be simulated with the present model as follows. Librarian ($A_e = 1.0$, $A_p = -1.2$, $A_a = -2.3$) and firefighter ($A_e = 2.1$, $A_p = 2.5$, $A_a = 1.8$) are used as examples for introvert versus extravert professional roles, respectively;[4] whereas to watch ($B_e = -0.3$, $B_p = -0.2$, $B_a = -1.8$) and to debate with ($B_e = 1.2$, $B_p = 1.7$, $B_a = 2.4$) serve as introvert versus extravert behaviors during the interview. Then, "a librarian watches an interviewer" and "a firefighter debates with an interviewer" are role-consistent behaviors. According to the present model, the resulting affective deflection values are $D = 3.5$ and $D = 4.5$, respectively. In contrast, descriptions of the corresponding role-violating behaviors "a librarian debates with an interviewer" and "a firefighter watches an interviewer" result in much higher deflection ($D = 19.4$ and $D = 14.9$, respectively). The latter are situations where observers can be expected to engage in trait attributions in order to restore consistent representations of the events.

What therefore are the modifying adjectives that could be added to the "job applicants" to solve the affective incoherence? INTERACT's trait suggestions for the librarian, chosen out of hundreds of words in the affective dictionary, are extremely extravert: adventurous and active; whereas the role-discrepant behavior of the firefighter can be explained by introvert characteristics such as self-conscious and lonesome.

Victim derogation. In the classic experiment by Jones et al. (1973), the amount to which a supposed victim of rape was devalued by the participants could be predicted by the initial social role of the "victim." The German INTERACT software, based on the present model of impression formation, can predict the main results of that experiment: (a) devaluing a victim of rape is an effective measure to reduce the affective incoherence inherent in such an event and (b) the higher the initial social status of the victim, the higher the extent to which the participants devalue her, because of the greater affective deflection that has to be resolved. "A man rapes a wife" (high status of victim) is an event that involves extreme affective incoherence ($D = 34.9$), whereas "a man rapes an ex-wife" (lower status of victim) still involves considerable deflection, but less of it ($D = 22.4$). INTERACT suggests the following trait attributions toward the victim as most probable: despicable and cowardly. It should be noted that the same attributions are suggested for both the high status and low status condition of the experiment. However, this can be explained with the limited number of words available in the affective dictionary. A larger sample of words might well result in more fine-tuned attribution suggestions. The important point is that the present model correctly predicts the higher need for devaluing the victim in the high status condition. The desire for affective consistency of linguistic event representations is a driving force for social sense making. The event "a violent man rapes a despicable woman" that involves trait attributions for both actor and object leaves the observer with only little unresolved deflection ($D = 7.3$).

Discussion

The three example applications of the model described in this article show that impression formation and attribution in social situations can be explained parsimoniously by an underlying mechanism of affective coherence in the linguistic representation of the situation. As the present research was based on material in the German language, the results add important evidence for the hypothesis that impression formation, as it is modeled within the framework of affect control theory (Heise, 1979, 2007; MacKinnon, 1994; Smith-Lovin & Heise, 1988), is a ubiquitous process. Not only is the evaluation–potency–activity structure of emotional meaning valid across cultures, as Osgood et al. (1975) and, more recently, Fontaine et al. (2007) have shown, but the same seems to be true for the affective dynamics in the processing of language.

The question can be raised as to whether some particularities of the present model in comparison with existing similar models in American English (Heise, 1991; Smith-Lovin, 1987) and Japanese (Smith et al., 1994) reflect German cultural traits.

Before an answer to that question is attempted, a few critical comments must be made concerning methodological issues, because valid examination of cross-cultural differences requires strict comparability of the employed methods. The crucial point here is the semantic configuration of the sample event descriptions that the model is based on. Despite all efforts to replicate the design from the Japanese study (Smith et al., 1994), the exactness of the replication may have been reduced owing to a variety of reasons, including the use of a rather old affective dictionary (Schneider, 1989) as a database for determining the configuration of actors, behaviors, and objects in the events. Also, it is by no means guaranteed that the same pattern of interactions within and across EPA dimensions emerges in all cultures. Thus, what proves to be a "balanced design" in one culture might still produce methodological artifacts in another. Not all the interaction effects found statistically significant can be interpreted as easily in light of existing social-psychological theories as $B_e O_e$, which has been explained as reflecting a need for evaluative balance. Why, for example, do potent acts that lively actors direct at potent targets seem nicer (.05 * $A_a B_p O_p$)? Such odd effects, although substantially contributing to the predictive power of the model, might result from the specific configuration of sample sentences rather than from cultural distinctions in impression formation processes. Therefore, interpreting the structure of the present model in terms of cultural traits requires a lot of caution and cannot be more than preliminary.

The contention that people strive for evaluative balance of their mental representations is one of the best established principles of social psychology (e.g., Heider, 1946; Newcomb, 1953; Osgood & Tannenbaum, 1955). In affect control theory's models of impression formation, evaluative balance is contained in the predictors $A_e B_e$ (good actors are expected to act nicely, bad actors are expected to act poorly), $B_e O_e$ (good objects are expected to be treated nicely, bad objects are expected to be treated poorly), and $A_e B_e O_e$ (good actors are expected to treat good objects nicely and bad objects poorly, etc.). Surprisingly, those three effects turned out to be considerably smaller (or even, for $A_e B_e O_e$, not significant at all) in the present German-language model than they are in the comparable U.S. and Japanese studies. For example, $B_e O_e$ regularly emerged as one of the most important predictors of postevent actor evaluations in numerous studies in English (Gollob, 1968; Heise, 1991, 2007; Smith-Lovin, 1987). In contrast, its contribution to the predictive power of the German A'_e equation reported in Table 1 is rather small with only a .04 coefficient as compared with .12 in American English (Heise, 1991) and .15 in Japanese (Smith et al., 1994).[5]

Some caution should be retained, but if it were true that evaluative balance were of much less importance to Germans as compared with Americans and Japanese, this would mean that the social-psychological basis of moral judgment is quite different in German culture. Take punishing criminals as an example, which is an obvious instance of $B_e O_e$ (treating a bad person poorly). Of course, Americans, Japanese, and Germans all punish criminals, but interestingly, the extent in so doing seems to vary with the magnitude of $B_e O_e$: The German per capita incarceration rate is among the lowest worldwide and about seven times as low as in the United States (Hartney, 2006).

Admittedly, the fact that Japan's is even slightly lower than Germany's weakens the strength of the argument; but as most U.S. states do, Japan allows the death penalty, which can be considered the severest of all punishments, and which, in contrast, has been opposed by a majority of the public in Germany for decades (Allensbach, 2002) and was abolished long ago.

The second particularity of the present model relates to the potency dimension. Among the most important predictors in many equations was B_pO_e, the interaction between behavior potency and object evaluation. For example, actors who direct strong actions at nice targets are perceived as considerably less nice ($-.13 * B_pO_e$ in the A'_e equation), all else being equal. This effect somewhat usurps the place that B_eO_e has in the American and Japanese models: rather than expecting good actors to be *nice* to nice targets, Germans expect good actors to refrain from exerting *power* toward nice targets. Apparently, the model of impression formation reflects the findings from Hofstede's (1984) extensive cross-cultural research program cited above regarding the lower power distance of Germany compared with the United States and Japan. This result is also consistent with Schneider's (2004) observation that Americans attach more positive feelings than Germans to concepts that denote authority, such as judge, president, or business manager. The present model provides a plausible explanation. Judges and presidents often act powerfully on targets, which, in a culture that rejects powerful actions, must eventually lead to an enduring less positive evaluation of those social roles.

Some theoretical concepts employed here deserve discussion as they relate to important challenges for affect control theory's potential to amount to a unifying theory of social interaction. The first issue is "culture." Talking about *the* German culture is oversimplified, and a more differentiated look is needed. Although it would hardly be feasible (as too resource consuming) to develop distinct complete models for different subgroups of society, specific hypotheses should nevertheless be addressed. Does, for example, the rejection of power hold true across different age groups? What impact does the respondents' socioeconomic status have? Answering such questions is important to account for the heterogeneity of modern complex cultures and might also shed light on the relations between (changing) values and the affective structure of the language.

"Coherence" is a second concept that might need a closer look, as there are instances where low affective deflection is neither necessary nor sufficient to make a statement semantically coherent. That a mother beats her daughter might also be understandable if the daughter is "rude" rather than "submissive." The event "A brother has sex with his sister" produces only little affective deflection according to the present model, yet it is definitely disturbing and unacceptable in German culture. Perhaps, a promising future approach consists in combining affect control theory with parallel constraint satisfaction models of emotion, thought, and action (e.g., Thagard, 2006). Then, the desire to reduce affective deflection constitutes but one (yet very powerful) constraint among others in the control of social interaction.

A further limitation of the presented research is that the sample applications were not systematically chosen and, therefore, do not offer much more than a plausibility demonstration. However, a more substantial approach for validating the model has been reported elsewhere (Schröder & Scholl, 2009): German students, playing the role of organizational leaders in a computer-simulated business environment, were given opportunities to select among different actions toward virtual employees. It could be shown that their choices were negatively related to the affective deflection inherent in those actions, as computed with the present model.

Some theoretical challenges notwithstanding, the research described in the present article demonstrates that the affective structure of the language seems to reflect very well-documented, basic principles of social psychology. Inferences and attributions in social situations as different as a mother interacting with her daughter, a psychologist interviewing a job applicant, and a man raping a woman could be predicted in a way that is very plausible in the first case and that, in the latter two cases, corresponds to the results of textbook classics in experimental social psychology. The model of impression formation and attribution is based on three premises, which possibly amount to basic laws of psychology: First, people are bound to the cognitive categories imposed by language when they ascribe meaning to social situations and base their actions thereupon. Second, evaluation, potency, and activity (sometimes labeled differently by different authors) are ubiquitous dimensions of social perception (Clore & Pappas, 2007; Mehrabian, 1980; Scholl, 2009) along which mental representations are organized. Third, people strive for coherence of their mental representations. These premises reflect the ever-present two-sided nature of human sociality: We are creatures of nature, who are ubiquitously bound to the same natural laws when controlling our experiences (e.g., maximizing coherence within a three-dimensional emotional space); at the same time, we are creatures of culture, who use specific symbolic systems to make sense of our social lives.

Notes

1. Following a research tradition in affect control theory, male and female ratings were treated separately for estimating the model. In a first step of the analysis, a dummy variable (0 = *male*, 1 = *female*) was created to control for potential sex-of-rater effects on the impression formation model. Additional interaction terms were created by multiplying the dummy variable with all the predictors. However, the inclusion of such sex-of-rater specific predictors barely contributed anything to the predictive power of the model. Also, the appearance of such effects seemed random based rather than systematic; hence, for reasons of parsimony, only one unisex model was estimated finally.

2. Again, male and female EPA ratings were treated separately, but a unisex model was estimated, as the procedure described in Note 2 (including a sex-of-rater dummy variable) did not result in substantially different male and female equations.

3. http://www.tschroeder.eu

4. In the original Jones et al. (1961) experiment, professional identities of astronaut and submariner were used as examples. However, it is not obvious exactly which kind of personality is required for those professions; hence, the subjects received a detailed explanation beforehand of why the ideal astronaut should have an introvert personality, whereas the position of submariner required extraversion. As the impression formation model presented here relies more on intuition, different examples were chosen.

5. Schröder and Heise (2008, unpublished data) tested the hypothesis that the smaller $B_e O_e$ effect in the German model occurred as a result of the semantic configuration of the 100 sample events rather than reflecting a cultural distinction in the impression formation process. They translated 20 additional actor–behavior–object event descriptions aimed at maximizing the variance of evaluative behavior–object balance into German that had been used in a previous study in the United States (Smith-Lovin, 1987). Examples for balanced events are "a landlady amuses a grandfather" and "a priest spanks a mugger," whereas phrases such as "a priest rescues a mugger" were used as unbalanced events. A sample of psychology undergraduates (19 male, 23 female) at the Humboldt University, Berlin, Germany, rated their in-context actor impressions with the semantic differential. Subsequently, the A'_e equation was estimated again, using the same procedure as described in the results section of this article but based on all 120 sample event ratings (100 from the large study plus the 20 $B_e O_e$ events). In the resulting new equation, the $B_e O_e$ coefficient turned out to be slightly higher (.07 instead of .04), but still considerably below the U.S. and Japanese standards.

References

Allensbach. (2002). *Allensbacher Jahrbuch der Demoskopie,* Band 11, 1998-2002 [Allensbach almanac of opinion surveys, Vol. 11, 1998-2002]. Munich, Germany: K. G. Saur.

Averett, C. P., & Heise, D. R. (1987). Modified social identities: Amalgamations, attributions, and emotions. *Journal of Mathematical Sociology, 13*, 103-132.

Clore, G. L., & Pappas, J. (2007). The affective regulation of social interaction. *Social Psychology Quarterly, 70*, 333-339.

Fontaine, J. R. J., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological Science, 18*, 1050-1057.

Gollob, H. F. (1968). Impression formation and word combination in sentences. *Journal of Personality and Social Psychology, 10*, 341-353.

Hartney, C. (2006). US rates of incarceration: A global perspective. *Fact sheet—research from the national council on crime and delinquency*. Retrieved from http://www.nccd-crc.org/nccd/pubs/2006nov_factsheet_incarceration.pdf

Heider, F. (1946). Attitudes and cognitive organization. *Journal of Psychology: Interdisciplinary and Applied, 21*, 107-112.

Heise, D. R. (1979). *Understanding events: Affect and the construction of social action*. New York, NY: Cambridge University Press.

Heise, D. R. (1991). *OLS equation estimations for INTERACT*. Retrieved from http://www.indiana.edu/~socpsy/ACT/references.html

Heise, D. R. (1997). *INTERACT. Introduction and software*. Retrieved from http://www.indiana.edu/~socpsy/ACT/interact.htm

Heise, D. R. (2001). Project Magellan: Collecting cross-cultural affective meanings via the internet. *Electronic Journal of Sociology, 5*. Retrieved from http://www.sociology.org

Heise, D. R. (2007). *Expressive order: Confirming sentiments in social action*. New York, NY: Springer.

Heise, D. R. (2010). *Surveying cultures. Discovering shared conceptions and sentiments*. Hoboken, NJ: Wiley.

Hofstede, G. (1984). *Culture's consequences: International differences in work-related values*. London: Sage.

Jacobs, I., & Scholl, W. (2005). Interpersonale Adjektivliste (IAL): Die empirische Umsetzung theoretischer Circumplex-Eigenschaften für die Messung interpersonaler Stile [Interpersonal adjective list (IAL): Measuring interpersonal styles based on the theory of the circumplex]. *Diagnostica, 51*, 145-155.

Jones, C., & Aronson, E. (1973). Attribution of fault to a rape victim as a function of respectability of the victim. *Journal of Personality and Social Psychology, 26*, 415-419.

Jones, E. E., & Davis, K. E. (1965). A theory of correspondent inferences: From acts to dispositions. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219-266). New York, NY: Academic Press.

Jones, E. E., Davis, K. E., & Gergen, K. J. (1961). Role playing variations and their informational value for person perception. *Journal of Abnormal and Social Psychology, 63*, 302-310.

Lienert, G. A., & Raatz, U. (1994). *Testaufbau und Testanalyse* (5. Aufl.) [Construction and analysis of psychological tests, 5th ed.]. Weinheim, Germany: Beltz.

MacKinnon, N. J. (1994). *Symbolic interactionism as affect control*. Albany: State University of New York Press.

McCrae, R. R., & Costa, P. T. (1989). The structure of interpersonal traits: Wiggins's circumplex and the five-factor model. *Journal of Personality and Social Psychology, 56*, 586-595.

Mehrabian, A. (1980). *Basic dimensions for a general psychological theory*. Cambridge, MA: Oelgeschlager, Gunn & Hain.

Mehrabian, A. (1995). Relationships among three general approaches to personality description. *Journal of Psychology: Interdisciplinary and Applied, 129*, 565-581.

Morgan, R. L., & Heise, D. (1988). Structure of emotions. *Social Psychology Quarterly, 51*, 19-31.

Newcomb, T. M. (1953). An approach to the study of communicative acts. *Psychological Review, 60*, 393-404.

Osgood, C. E., May, W. H., & Miron, M. S. (1975). *Cross-cultural universals of affective meaning*. Urbana: University of Illinois Press.

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

Osgood, C. E., & Tannenbaum, P. H. (1955). The principle of congruity in the prediction of attitude change. *Psychological Review, 62*, 42-55.

Romney, A. K., Weller, S. C., & Batchelder, W. H. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist, 88*, 313-338.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information, 44*, 693-727.

Schneider, A. (1989). *Mean affective ratings of 804 concepts by West German students in 1989* [computer file]. Distributed at Affect Control Theory Website, Program Interact. Retrieved from http://www.indiana.edu/~socpsy/ACT/interact/JavaInteract.html

Schneider, A. (2004). The ideal type of authority in the United States and Germany. *Sociological Perspectives, 47*, 313-327.

Scholl, W. (2009). The basic socio-emotional dimensions of human cognition, communication and interaction. In W. Scholl (Ed.), *Interact communicate: An interdisciplinary Wiki on social interaction and communication*. Retrieved from http://www.socialinteractionresearch.com

Schröder, T., & Scholl, W. (2009). Affective dynamics of leadership: An experimental test of affect control theory. *Social Psychology Quarterly, 72*, 180-197.

Skrandies, W. (1998). Evoked potential correlates of semantic meaning: A brain mapping study. *Cognitive Brain Research, 6*, 173-183.

Smith, H. W., Matsuno, T., & Ike, S. (2001). The affective basis of attributional processes among Japanese and Americans. *Social Psychology Quarterly, 64*, 180-194.

Smith, H. W., Matsuno, T., & Umino, M. (1994). How similar are impression formation processes among Japanese and Americans? *Social Psychology Quarterly, 57*, 124-139.

Smith-Lovin, L. (1987). Impressions from events. *Journal of Mathematical Sociology, 13*, 35-70.

Smith-Lovin, L., & Heise, D. R. (1988). *Analyzing social interaction: Advances in affect control theory*. New York, NY: Gordon and Breach Science.

Thagard, P. (2006). *Hot thought:* Mechanisms and applications of emotional cognition. Cambridge: MIT Press.

Universität Leipzig. (2007). *Portal deutscher Wortschatz* [German thesaurus Internet site]. Retrieved from http://wortschatz.uni-leipzig.de

Wish, M., Deutsch, M., & Kaplan, S. J. (1976). Perceived dimensions of interpersonal relations. *Journal of Personality and Social Psychology, 33*, 409-420.

Bio

Tobias Schröder is a research associate at the Freie Universität Berlin. The present article is based on his PhD dissertation thesis submitted at the Humboldt-Universität zu Berlin in 2009. His research interests include the connections between language, emotions, and social interaction, as well as the social-psychological basis of political decision making.