Empirically Investigating Determinants of Normative Processes: A Comparison of Two Approaches

David R. Heise

Department of Sociology

Indiana University

Bloomington, IN

Abstract

Two methods were applied to discover determinants of normative sentiment changes. Analyses of variance identified credible predictors somewhat more often than did stepwise regressions. With both methods, results from sub-samples constituted approximate subsets of results from larger samples, indicating that key effects can be found in small studies, but not all effects. About three-quarters of the specifications obtained with one method also were obtained with the other method. However, unique specifications from analyses of variance were more credible than those from stepwise regressions. The advantage of analyses of variance over stepwise regressions derived from dichotomizing exogenous variables, which addressed three problems in data on normative sentiment change: constricted variances, multicolinearity, and excessive influence of a few extreme cases. Structural equation estimations using specifications from analyses of variance showed that group differences—females versus males in this case—occur in size of coefficients as well as in variant specifications.

Introduction

Norms and normative processes permeate social life, yet have received less methodological attention than the study of attitudinal and behavioral heterogeneity. Methodologically, the study of norms departs from studies of heterogeneity in several ways. For instance, normative studies require different survey methods than the sample surveys that are required for studying heterogeneity (Heise 2010; Rossi and Nock 1982). Another difference is that assessing norms is more like a problem of scale creation with informants as indicators of a norm than a matter of measuring the central tendency of a distribution (Heise 2010). Additionally, as examined in this article, causal analysis and structural equation modeling of normative processes often require empirical specification of non-linear interactions among exogenous variables.

One notable methodological tradition contributing to principled study of norms and normative processes is the "factorial surveys" line of research, initially systematized in Rossi and Nock's (1982) book on the topic. In the last three decades, scores of studies have used factorial surveys, as reviewed by Wallander (2009). Jasso (2006) provided a unified framework for factorial surveys relating to normative beliefs and judgments.

Another tradition has its roots in the "culture as concensus" approach in anthropology (Romney 1994; Romney, Batchelder, and Weller 1987; Romney, Weller, and Batchelder 1986). Heise (2010) elaborated this methodological perspective sociologically, giving special attention to the topic of normative sentiments and normative sentiment change resulting from social actions. An example from each of these perspectives illustrates some issues that arise in studying normative processes.

The factorial survey approach is illustrated by a study of responses to a family member's emergent life conditions. Rossi and Rossi (1990) developed 1,628 vignettes in which an individual was experiencing either a crisis or a celebratory occasion. Familial relationships between the vignettes' protagonists and the respondent varied from nuclear family through extended family to non-familial relations like neighbor. The vignettes also varied the kinds of crisis or celebration being experienced by the protagonist. Respondents rated the kinds of help or recognition they might offer in terms of financial aid, emotional support, gifts, or visiting. Each respondent in the survey dealt with just 31 of the 1,628 vignettes, the judgment task being spread out over the full sample, with about 22 respondents providing data on any one vignette , and ratings by different respondents of different vignettes were pooled in order to analyze all vignettes together. The presumption was that large samples of respondents were not required for each judgment because matters of consensus rather than diversity were at issue. Through regression analyses and comparisons of mean ratings, the Rossi's determined that normative variations in help-giving and approbation among relatives primarily are determined by closeness of relationship, though some other factors like the sex and marital status of the protagonist also have effects.

The culture-as-consensus approach is illustrated by cross-cultural research on how a social action normatively transforms individuals' sentiments about a protagonist from one state to another (Gollob 1968; Gollob and Rossman 1973; MacKinnon 1985; Schröder 2011; Smith, Matsuno, and Umino 1994; Smith-Lovin 1987). Such impression-formation studies are the empirical basis of affect control theory (Heise 2007; Hoey, Schröder, and Alhothali 2013; MacKinnon 1994; Robinson and Smith-Lovin 2006), which mathematically predicts role behaviors, identities, attributions, and emotions, given verbal definitions of social situations and preceding actions. In these studies, respondents are presented with terse vignettes describing actions, and they use quantitative scales in order to rate how they feel about the actor, behavior, and object of behavior as a result of the action. The same scales are used to measure pre-event feelings about each of the elements in each event. Feelings are measured on three different quantitative scales, so in total each vignette requires eighteen measurements assessing feelings before and after the action. Empirically derived equations describe the transformations generated by the event, and include interaction terms to predict phenomena such as an actor seeming acceptable despite engaging in a cruel behavior, providing the behavior was directed at an evil person so that behavior and object are consistent in evaluation.

This article examines two empirical methods for specifying causal determinants of normative processes, using a legacy dataset from impression formation research. Impression-formation research provides an especially useful basis for comparing methods because it offers a meta-theory allowing empirical results to be judged more or less likely as possible representations of reality. Additionally the available dataset on impression formation is large enough to examine how well smaller datasets reproduce results from the larger dataset. Knowledge about causal factors also is used as a specification framework for estimating structural models of normative change in different cultures or groups.

Assessing Normative Processes

Rossi and Anderson (1982) argued that observations of actual behavior are not the best way to assess normative processes, because key cases may occur so rarely as to frustrate the progress of research. For example, a study of class and sex effects on judges' sentencing decisions rarely would encounter critical cases like the sentencing of middle-class burglars or college graduate muggers or female rapists in a typical run of court cases. Instead of observing actual behaviors, Rossi and Anderson proposed using vignettes in order to see how critical variations in descriptions of situations influence individuals' reactions along scales of response, such as length of prison sentence.

Rossi and Anderson referred to systematically constructed vignettes as factorial objects. Factorial objects are generated by combining multiple factors, each factor having multiple levels, with a level of a factor constituting one of the characteristics that is incorporated into a vignette. A set of vignettes that systematically varies factor levels and their combinations allows one to determine the extent to which each factor contributes to response variation.

Ideally one would specify the causal factors affecting a normative process theoretically in order to regress responses on determinants and thereby quantify the impacts of different causes. However, no general theory offers hypotheses regarding the determinants of many normative processes because causal processes involved in normative processes are cultural facts at least as much as they are components of some overall theory. For example, behavior-object evaluative consistency contributes to evaluation of an actor in the U.S.A. (Gollob 1968; Smith-Lovin 1987), it also affects evaluation of an actor in Germany but with only half the strength of the U.S.A. effect (Schröder 2011), and it has no effect on actor evaluation among Canadians according to an analysis[1] using the findings of this article. Thus predicting effects of behavior-object evaluative consistency within a particular culture is beyond prevailing theory, and guidance must be sought empirically concerning the causes of normative sentiment change in that culture. Indeed, a primary motivation for conducting a study is to discover the determinants of a normative process within a particular culture.

Quantitative studies of normative processes typically have dual functions: to uncover the determinants of a normative change, and to estimate coefficients describing how these causes quantitatively impact outcomes. The first function requires determining what causal mechanisms are involved in the process within the given culture—the specification problem. The second function requires estimating the quantitative impact of each determinant in the form of coefficients for equations that represent those processes—the estimation problem. Before equations can be estimated, their terms have to be specified.

The dual goals of model specification and equation estimation sometimes are accomplished simultaneously via stepwise regressions, notwithstanding the cautions expressed by statisticians like Cook and Weisberg (1999:280). Stepwise regressions can be conducted in two ways. Backwards stepping starts with the dependent variable regressed on all predictors, and removes variables with insignificant effects, until only significant regression coefficients are left[2]. Forward stepping starts with a constant, and adds variables with significant effects until no more significant regression coefficients can be found. In both cases, variables are removed or added one at a time, with the regression being re-computed after each step.

An alternative method for assessing the causal determinants of a normative process is available in some cases. Vignette studies of normative processes usually can be viewed as experiments, and determining which factors and their interactions impact subjects' responses can be accomplished through analyses of variance. While analysis of variance is a version of the general linear model, just as regression analysis is, there is an important difference: the factors of an analysis of variance are categorical instead of continuous. Converting quantitatively measured pre-conditions into categorical factors gives all cases equal influence, and ameliorates problems of differences in the variances of the determinants, as well as problems of correlated predictors.

Impression Formation Research

Impression formation research addresses questions like the following. How does a person's action towards someone make the actor seem attractive or unattractive, helpful or unhelpful, virtuous or evil? What is it about being the object of another's action that makes a person seem weak, ineffective, and irrelevant, as opposed to powerful, efficacious, and important? How does the morality of an action change depending on circumstances—e.g., even homicide is less bad if the action is directed at an attacking maniac? Well-done impression-formation studies provide rigorous, quantitatively buttressed answers to such questions.

Studies in more than a score of societies established three dimensions of affective response (Heise 2010; Osgood, May, and Miron 1975). *Evaluation* (valence) contrasts good versus bad; *Potency* (control) contrasts powerful versus powerless; and *Activity* (arousal) contrasts activated versus quiet. Stimuli such as smells, sounds, concepts, kinds of people, and elements of action can be measured on these dimensions with bipolar rating scales anchored on either end by contrasting adjectives. The measured affect associated with a stimulus is a vector in Evaluation-Potency-Activity (*epa*) space, with Cartesian coordinates designated by the ordered triplet of *epa* values. Heise (1969; 2010) discussed various issues in measuring *epa*.

Impression-formation processes are studied with vignettes verbally describing actions, and *epa* measurements of the actor (A), behavior, (B) and object of behavior (O) are obtained before and after each action. The goal is to create a set of equations characterizing how an action transforms affective impressions existing before the action (A_e, A_p, A_a, B_e, B_p, B_a, O_e, O_p, O_a) into new impressions after the action. Because transformations are assumed uniform for all kinds of actors, behaviors, and objects, no interactions relate to the same action element. For example, the interaction between actor evaluation and actor potency is not considered, nor are very high-order interactions like $A_e{\cdot}A_p{\cdot}B_e{\cdot}B_p{\cdot}O_p$ because they interrelate *epa* values of the same action element. Accordingly just 64 possible predictors are of interest in impression-formation studies: a constant, nine first-order terms representing impressions before the action (e.g., A_e, A_p, A_a, B_e), 27 second-order interaction terms (e.g., $A_e{\cdot}B_e$, $A_p{\cdot}B_e$), and 27 third-order interaction terms (e.g., $A_e{\cdot}B_e{\cdot}O_e$, $A_e{\cdot}B_a{\cdot}O_p$). The full transformation from pre-action to post-action sentiments is delineated in nine such equations, each consisting of multiple terms predicting one of the outcome impressions (e.g., A_e', A_p', A_a', B_e').

Let an action be categorized by indicating positive versus negative *epa* values associated with the actor, behavior, and object. For example, "an athlete is entertaining a teammate" is of type A_{e+p+a+} B_{e+p+a+} O_{e+p+a+}; "a psychotic is confining an athlete" is of type A_{e-p-a+} B_{e-p+a-} O_{e+p+a+}. There are 512 such types of action, and a full-factorial design for an impression-formation study ideally would include all of them, in order to cross all plus and minus *epa* profiles of actors, behaviors, and objects with each other. However, such a study has been attempted only once (Smith-Lovin 1987) because of the large amounts of data that have to be collected. To fulfill the design and meet minimal standards of norm measurement, each of the 512 actors, behaviors, and objects has to be rated on the three *epa* dimensions both out of context and in the context of the specific action, and ratings must be done by at least 20 respondents in order to assess normative views. Thus the full-factorial design requires more than 180,000 measurements. One way to economize is to use the same identity in multiple actions, such as the repetition of "athlete" in the examples above. Additionally, most impression-formation studies (Gollob 1968; MacKinnon 1985; Schröder 2011; Smith, Matsuno, and Umino 1994) have economized by dealing with only 100 to 214 action types. Actions in some of these studies were selected systematically and in other studies were sampled randomly from the 512 types. Heise (2010) proposed a systematically condensed design of 128 actions for future work. The Heise design has 64 action sentences with evaluation and potency fully crossed and another 64 action sentences with potency and activity fully crossed, the third dimension being varied as a Latin square in each case. In this design every two-way interaction has each of its possible variations represented by 32 different action sentences, while 16 action sentences represent each variation of every three-way interaction.

A key goal of impression-formation research is to identify complex processes that affect normative sentiment change, like the interaction of behavior evaluation and object evaluation that condones an actor's cruelty toward abhorrent individuals. From the standpoint of social psychological theory, interaction terms that involve a single dimension are easier to interpret than terms that mix two or three affective dimensions. For example, actor evaluation being affected by the $B_e{\cdot}O_e$ interaction is easy to interpret as a consistency effect with theoretical provenance tracing back to Heider (1958). Acceptance of an actor who engages in cruel behaviors (B_e negative) toward an evil person (O_e negative) can be represented by multiplying the two negatives and producing a positive value which is incorporated into actor evaluation.

Terms that mix two dimensions are somewhat harder to interpret, but they also are known to be operative. For example, $B_e{\cdot}O_p$ with a negative coefficient has an established theoretical interpretation (Gollob and Rossman 1973), has been replicated cross-culturally (Smith-Lovin 1987; Smith, Matsuno, and Umino 1994), and functions to integrate displays of mercifulness, courageousness, sycophancy, and ruthlessness into actor evaluation (Heise 2007).

No theoretical interpretations have been developed yet for effects that mix all three affective dimensions, but some such effects have been found. For example, in Japan (Smith, Matsuno, and Umino 1994) the impression of an actor's potency is influenced negatively by the actor's original activation interacting with evaluation of the actor's behavior, $A_a{\cdot}B_e$. To illustrate, an active actor like a boy loses potency (seems a sissy) when acting too nicely. A second-order interaction like this may be intuitively comprehensible, but a third-order interaction involving all three dimensions can seem beyond comprehension and theoretical explanation. For example, Schröder (2011:15), who found several such terms in his study of Germans' impression formation, forwent interpreting them and instead concluded, "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."

Overall, the less an equation term mixes *epa* dimensions, the easier it is to interpret and encompass theoretically, so let us classify a term as *credible* if all its components are on the same dimension as the dependent variable; as *feasible* if the term and the variable it predicts involve two dimensions; and as *implausible* if the term and the variable it predicts involve three dimensions. Note that implausible does not mean impossible—only that such an effect is difficult to theorize in contemporary social psychology and thereby seems relatively unlikely. Notwithstanding the unlikeliness, one typically wants to investigate empirically the possibility that such factors may be at work in a particular culture. Even terms that are irrelevant in a current study should be re-examined in future studies, because processes may vary from culture to culture, generation to generation, or group to group.

Data

The largest impression-formation study (Smith-Lovin 1987) dealt with 515 actions, 512 of which were intended to constitute a full-factorial design, and three of which were intended to represent actions with neutral actors, behaviors, and objects[3]. Out-of-context ratings were obtained from 55 respondents, and in-context ratings were obtained from 44 respondents, with numbers of male and female respondents about equal. In the overall study (which included several components besides impression formation), 1,225 North Carolina undergraduates participated, and

5,345 stimuli were presented on 40 different paper forms. Ratings were made on nine-position rating scales, the positions being assigned quantitative values via successive-intervals scaling before computing the mean *epa* profile for each stimulus. The evaluation rating scale was anchored with the adjectives *good, nice* versus *bad, awful*; the potency scale with *big, powerful* versus *little, powerless*; and the activity scale with *fast, young, noisy* versus *slow, old, quiet*. Smith-Lovin (1987) provided additional details concerning data collection and processing.

The factorial design did not materialize entirely when patterns were identified from actual mean out-of-context ratings obtained in the study. Instead of having 513 distinct patterns there were only 267, based on male out-of-context means, and 195 when patterns were defined with female means. Moreover, stimuli were not distributed evenly into various actor-behavior-object configurations on Activity and Potency. In particular, action patterns involving weak behaviors ($p < 0$), such as $A_{e.p+a.}B_{e.p\text{-}a.}O_{e.p\text{-}a.}$, constituted between three and seven percent of the 515 sentences, though each such pattern should have contributed twelve percent. Similarly, Activity patterns involving quiet behaviors ($a < 0$) contributed between three and five percent of the 515 action sentences, instead of twelve percent. These problems actually are endemic in impression-formation research in that most interpersonal behaviors are not weak or quiet. Only eight percent of the verbs rated in the study had negative potencies, and none were more than slightly weak according to combined male-female means. Similarly, just nine percent of the verbs had negative activities, with none more than slightly quiet.

Notwithstanding the breakdown in factorial design, the data on 515 actions provide the optimal existing basis for impression-formation analyses, and these are the data used here to examine methodological approaches to specification and estimation of prediction equations. The failure to achieve a perfect design actually is a benefit for methodological analyses, providing a messy empirical reality as opposed to a mathematical idealization. In a perfect factorial implementation, the first-order predictor variables correlate zero with one another, and any variables constructed by multiplying the first-order variables also correlate zero with one another. However, in the 515 dataset, the three behavior variables—B_e, B_p, B_a—correlate from 0.14 to 0.31 for males, and from 0.05 to 0.36 for females. These and other first-order correlations propagate eleven correlations of 0.30 or more in the second-order terms for males, and 19 for females, plus twelve correlations of 0.30 or more in the third-order terms for males, and 18 for females.

A colinearity problem arises when correlations among predictors are substantially greater than correlations between predictors and dependent variables. This was not an issue among the first-order variables where every dependent variable had a predictor correlating 0.62 or more for males, and 0.60 or more for females. For correlations between second-order predictors and dependent variables the figures were 0.29 for males and 0.33 for females, so colinearity still was not very problematic. However, maximum correlations between third-order predictors and dependent variables started at 0.12 for males, and 0.14 for females, values that are far below some of the correlations among third-order variables, and therefore the 515-actions dataset presents realistic colinearity issues with regard to third-order terms.

Besides being used in its entirety, the dataset on 515 actions was used to generate sub-datasets in order to test the supposition that statistical results based on small samples of actions contain major effects found in a large sample of actions, while lacking some lesser effects because of the lower statistical power of the small sample. In other words, small sample results might be a subset of large sample results. Using Heise’s (2010) design as a guide, actions were extracted from the 515-actions dataset, once based on male data and once based on female data.

Patterns in the 515 dataset should have contained all 128 action types in the Heise design, but they did not because of the problems in actualizing the original factorial design. No match was available for 47 (37 percent) of the 128 patterns in the Heise design when defining patterns with male means, and for 45 (35 percent) of the 128 patterns in the Heise design when defining patterns with female means. Imperfect matches were substituted for the ideal patterns in most of these cases, allowing a low positive value (usually less than 1.0) to serve in place of a designated negative value. With such compromises, the male subset comprised 122 patterns, and the female subset 117 patterns, as compared to the 128 patterns in the ideal Heise design. Each sub-dataset was filled out to 128 actions with patterns close to the missing ones.

Henceforth the data for 515 actions is referred to as the *515-actions dataset*, and the data for 128 actions is the *128-actions dataset*.

Stepwise Regressions

Using the 515-actions female and male datasets, backward and forward stepwise regressions were conducted for each of the nine predicted variables. The backward stepwise regressions removed terms with insignificant coefficients, $p >= .01$, and the forward stepwise regressions added terms with significant coefficients, $p < .01$ (and then removed them if they turned insignificant in later steps, $p >= .05$). Results are summarized in the top segment of Table 1.

Backward and forward stepwise regressions also were conducted with the 128-actions female and male datasets, and results from 128-actions stepwise regression analyses were compared with results from 515-actions stepwise analyses in order to assess what is lost by using stepwise regression on relatively small samples of actions. The analyses of 128-actions datasets used the same removal and addition criteria as the analyses with 515 actions. Results from these analyses are summarized in the bottom part of Table 1.

Females

Backward stepwise analysis of the 515-actions female dataset produced 70 predictors in nine equations (involving 26 different variables and interactions, out of 63 possible terms). Forward stepwise analysis yielded 69 predictors in the nine equations (involving 25 different terms).

Over the nine equations, the two methods generated equations having 63 predictors in common (23 different terms), of which 28 were credible, 31 feasible, and four implausible. Backwards stepping produced seven terms that did not emerge in forward stepping, and forward stepping produced six terms that were not in the equations from backward stepping. Backward stepping produced more terms that were difficult to interpret (feasible or implausible rather than credible) than did forward stepping.

Backward analyses of 128-actions yielded 35 predictors in nine equations (18 different terms). Twenty-seven of the 35 predictors were the same as were obtained by backward analysis of the 515-actions dataset, and 20 of these were credible, seven feasible, and none implausible. One of the twelve terms emerging in the analyses of 128-actions that did not emerge with 515-actions one was credible, seven were feasible, and none were implausible.

Forward analyses of 128-actions yielded 31 predictors in nine equations (16 different terms). Twenty-five of the 31 predictors were the same as were obtained by forward analysis of the 515-actions dataset, and 19 of these were credible, six feasible, and none implausible. Two of the seven terms emerging in the analyses of 128-actions that did not emerge with 515-actions were credible, three were feasible, and one was implausible.

The conjunction of backwards and forward results with 128 actions had 29 predictors, 20 of which were credible, and the rest feasible. Twenty-five of the predictors were the same as appeared in the conjunctive results for 515 actions, and 19 of these were credible with the rest feasible. Four terms in the conjunction of results for 125 actions were not present in the conjunction of results for 515 actions; one of these was credible, three feasible.

Table 1								
Number and Interpretability of Effects Specified Via Stepwise Regressions of Impression-Formatior								
	Females	Males						
	Credible	Feasible	Implausibl	Total	Credible	Feasible	Implausibl	Total
515 Actions								
Same in Bc	28	31	4	63	27	32	1	60
Unique to	1	5	1	7	2	9	1	12
Unique to	3	3	0	6	0	3	1	4
128 Actions								
Same in Bc	20	9	0	29	23	15	4	42
Unique to	1	5	0	6	1	7	0	8
Unique to	1	0	1	2	0	0	0	0
Same in 51	28	7	0	35	23	15	0	38
Unique to	1	7	0	8	1	7	4	12
Same in 51	19	6	0	25	23	12	0	35
Unique to	2	3	1	6	0	3	4	7
Same in 51	19	6	0	25	23	11	0	34
Unique to	1	3	0	4	0	4	4	8

Males

Backward stepwise analysis of the 515-actions male dataset ended up with 72 predictors in nine equations (involving 28 different terms). Twenty-nine of the predictors were credible, 41 feasible, and two implausible. Forward stepwise analysis yielded 64 predictors in the nine equations (involving 24 different terms), and 27 of these were credible, 35 were feasible, and two were implausible.

The two methods generated equations having 60 predictors in common (21 different terms), of which 27 were credible, 32 feasible, and one implausible. Backwards stepping produced twelve terms that did not emerge in forward stepping, and forward stepping produced four terms that were not in the equations from backward stepping. Most of the extra effects identified by backward stepping were relatively difficult to interpret (feasible or implausible rather than credible).

Backward analyses of 128-actions yielded 50 predictors in nine equations (23 different terms). Thirty-eight of the 50 predictors were the same as were obtained by backward analysis of the 515-actions dataset, and 23 of these were credible, 15 feasible, and none implausible. One of the twelve terms emerging in the analyses of 128-actions that did not emerge with 515-actions was credible, seven were feasible, and four were implausible.

Forward analyses of 128-actions yielded 42 predictors in nine equations (20 different terms). Thirty-five of the 42 predictors were the same as were obtained by forward analysis of the 515-actions dataset, and 23 of these were credible, 12 feasible, and none implausible. None of the seven terms emerging in the analyses of 128-actions that did not emerge with 515-actions were credible, three were feasible, and four were implausible.

The conjunction of backwards and forward results with 128 actions had 42 predictors, 23 of which were credible, 15 were feasible, and four implausible. Thirty-four of the predictors were the same as appeared in the conjunctive results for 515 actions, and 25 of these were credible with the rest feasible. Eight terms in the conjunction of results for 125 actions were not present in the conjunction of results for 515 actions; four of these were credible, and four were implausible.

Summary

Pooling results from female and male analyses and analyses of nine dependent variables gives a base of 18 specification problems for examining the usefulness of stepwise regression in identifying causal determinants of impression formation.

With the 515-actions dataset, 40 percent of the terms identified in either backward or forward stepwise regressions were easily interpretable (in the credible category); the figure was 53 percent with the 128-actions data. The remaining terms mostly were in the feasible category, though at least five percent of all identified terms were in the difficult-to-interpret implausible category.

Retaining only terms that appeared conjunctively in both backward and forward analyses provided marginally better results: 45, 51, and four percent were in the credible, feasible, and implausible categories when analyzing 515 actions, and 61, 34, and six percent when analyzing 128 actions. Only conjunctive results are considered below.

Most effects found with the 128-actions subsample also were effects found with the full sample (83 percent). However, the effects found uniquely in the subsample mostly were complex multidimensional interactions (92 percent), and a third of these were implausible terms mixing all three *epa* dimensions. Thus, analyses with the subsample identified effects found with the larger sample, but also brought in a number of chaff terms, including some that would be difficult to interpret theoretically.

An attempt was made to improve on stepwise regression results by employing a more modern automatic selection algorithm—Least Angle Regression or LARS (Efron, Hastie, Johnstone, and Tibshirani 2004), with number of retained predictors corresponding, as recommended, to the minimum value of C_p (a measure of model fit). However, LARS results were more disorderly than stepwise regression results in this particular problem. LARS analysis of the female 515-actions dataset identified 61 of the 63 available predictors as determinants in one or more of the nine equations, while LARS analyses of the male 515-actions dataset incorporated 62 of the 63 available predictors in one or more equations. Thereby, for both sexes, LARS incorporated dozens of implausible predictors into its suggested models of the nine-equation system.

Analyses of Variance

The impression-formation data were subjected to analysis of variance by coding each action as zero or one on nine factors: A_e, A_p, A_a, B_e, B_p, B_a, O_e, O_p, and O_a. Each of the nine factors was dichotomized by splitting at the median, within each sample considered—515-actions, 128-actions, and 192-actions (discussed below), for females and for males. Thereby half of all actions within each sample were categorized as, for instance, A_e-high and half as A_e-low; half were categorized as B_p-high and half as B_p-low; and so on.

The interactions that are of interest in the ANOVAs combine a factor relating to one action element with a factor relating to another action element—that is, the 54 interactions indicated previously. Amounts of variance in dependent variables explained by each of the specified factors and factor interactions were tested for significance, using the aggregate of residual variance and variance contributed by all unspecified interactions as error variance, and with the significance level set at the 0.01 level. Results of the analyses of variance are summarized in Table 2.

Table 2

Significant Terms From Analyses of Variance of Nine Dependent Variables

	Female					Male		
Sample Size	Credible	Feasible	Implausible	Total	Same as 515	Credible	Feasible	Implausible
515	28	27	0	55	…	28	29	0
128	18	6	0	24	23	19	8	0
192	21	10	1	32	29	23	10	1

Analyses of variance of the 515-actions female dataset found 55 predictors in nine equations (involving 19 different terms). Twenty-eight of the predictors were credible, 27 feasible, and none implausible. Analyses of the 515-actions male dataset produced 57 predictors (involving 19 different terms), of which 28 were credible, 29 feasible, and none implausible.

The 128-actions subsamples yielded fewer predictors—46 percent of the number found with the 515-actions dataset (sexes pooled). In the case of female data, there were 24 significant predictors in all nine equations (involving 11 different terms), and 18 of these were credible, six feasible, and none implausible. For males, there were 27 predictors (involving 14 different terms), with 19 of these being credible, eight feasible, and none implausible. Most of the predictors found with the subsample also had been found with the 515-actions sample: 96 percent in the case of females, and 89 percent in the case of males. In the case of females, the subsample gave one feasible effect not found with the 515-actions sample. In the case of males, the subsample had three feasible effects not found with the 515-actions sample.

A larger subsample of 192 actions was constructed in order to explore the impact of sample size. Analyses with the 192-actions sample address the question: How much better are results when the subsample size is increased by fifty percent? Sixty-four new action patterns were defined in which evaluation and activity were fully crossed and potency was incorporated through a Latin square[4]. The 64 extra patterns repeated none of the patterns in the 128-actions sample, so the two sets were combined to create a set of 192 action patterns. As before, a majority of the new patterns could be matched to action sentences in the 515-actions sample, but some patterns had to be implemented with sentences that matched the ideal imperfectly.

Considering female results with the 192-actions subsample, there were 32 significant predictors in all nine equations (involving 13 different terms), and 21 of these were credible, ten feasible, and one implausible. For males, there were 34 predictors (involving 14 different terms), with 23 of these being credible, ten feasible, and one implausible. Percentages of predictors found with the subsample that also had been found with corresponding 515-actions sample were 91 percent in the case of females, and 82 percent in the case of males. In the case of females, the subsample gave two feasible effects and one implausible effect not found with the 515-actions sample. In the case of males, the subsample had five feasible effects and one implausible effect not found with the 515-actions sample.

The 192-actions subsamples missed fewer of the predictors found with the 515-actions sample: 49 percent, as compared to the 128-actions subsample which missed 57 percent of the predictors found with the 515-actions sample (pooled female and male results in both cases). The 192-actions subsamples came up with more chaff than the 128-actions subsamples, though this probably signals that analyses of the 128-actions were free of chaff only by chance.

Parameter Estimation

Results from the analyses of variance specify causal processes in impression formation, and the specifications were used to estimate coefficients in structural equation models. Structural equation modeling returned to the *epa* values of predictors (rather than dichotomizations), and products of the first order predictors like B_e*O_e implemented interactions. Structural-equation estimations were set up within a seemingly-unrelated-regressions framework in which all nine impression-formation equations were estimated simultaneously.

Female and male causal determinants of each dependent variable were pooled in order to examine sex differences in coefficients. In cases of predictors that were unique to one sex, the coefficient for the other sex was constrained to zero. Constants in all equations were allowed to vary freely across sex.

Equations estimated with the 515-actions data used specifications from analyses of variance of the same dataset. Sixty-nine predictors were involved in the nine equations, plus constants. Forty-one of the predictors had been identified as causal determinants of impressions in both female and male analyses of variance, twelve predictors were unique to females, and 16 were unique to males. Coefficients for the predictors that were shared across the sexes were constrained to equality in one analysis, and freed in another analysis in order to test whether females and males weighted their shared impression-formation mechanisms the same. Equality constraints greatly increased the chi-square (d.f. = 41, chi-square difference = 136.24). Thus, even impression-formation processes that the sexes shared contributed to impressions differently for females and males. Wald tests for group invariance of parameters indicated that six coefficients in particular were significantly different across the sexes, $p < .01$.

Effects of small samples on variance explained were examined by estimating equations with the 128-actions data and specifications from the corresponding analysis of variance. Fifty-one predictors were involved in the nine equations, plus constants. Coefficients for the predictors that were shared across the sexes, and constants in all equations, were allowed to vary freely across sex. If a predictor was unique to one sex, the coefficient for the other sex was constrained to zero.

Explained variance varied from 58 to 81 percent over the nine equations in the small sample analysis. Explained variance varied from 61 to 84 percent over the parallel analyses with the 515-actions dataset. Thus, equations derived with small samples of actions identify the main contributors to impression formation processes, and the additional predictors obtained with larger datasets are relatively subtle effects influencing impressions only to small degrees, or only in unusual circumstances.

Discussion

Two questions were at issue in this study of alternative frameworks for establishing underlying causes of a normative process. First, how interpretable are results? The interpretability of results could be evaluated because the particular normative process being examined offered a meta-theory for assessing the plausibility of various predictive terms. The goal was to determine which framework yielded more credible results theoretically. Second, how well do results from smaller studies betoken results from larger studies? Female and male 128-actions datasets were extracted from the 515-actions datasets in order to address this question, and the subsamples, like the larger datasets, were carefully constructed to represent all possible causal factors. Results from the smaller dataset were examined to determine the extent that they constituted a subset of large study results, identifying the most important causal determinants found with the larger dataset, without bringing in extraneous predictors that were absent in results from the larger dataset.

The first analytic framework, stepwise regression, was tried in both the backward and forward variants. Forward stepwise regressions in which terms were added to a minimal model produced different results than backwards stepwise regressions in which terms were deleted from a maximal model. A core of predictors emerged in both variants—the conjunctive set of predictors, and the conjunctive set of stepwise results was used in some of the comparisons here. However, there also were a large number of predictors emerging with one stepwise variant but not the other, and some of these terms were implausible because of their complexity.

The second analytic framework was analysis of variance. Each of the basic predictor variables assessing pre-event feelings about an action's actor, behavior, or object (A_e, A_p, A_a, B_e, B_p, B_a, O_e, O_p, O_a) was split at the median in order to define nine first-order dichotomous factors with equal variance. The interactions tested for significance were those that crossed factors relating to different action components (e.g., $B_e x O_e$ or $A_e x B_e x O_e$), while variance from interactions relating multiple factors of the same component (e.g., $A_e x A_p x A_a$) was incorporated into the error term.

The two analytic frameworks converged in their specifications to some degree. Seventy-seven percent of the predictors found with analysis of variance showed up among the predictors found in the conjunctive stepwise results; and 70 percent of the stepwise predictors appeared among those found with analysis of variance (pooling results from all nine equations over both sexes). However, each analytic framework also came up with some unique predictors. Analyses of variance identified 26 predictors that were passed over in stepwise results, of which five were credible and 21 were feasible. Stepwise analyses found 37 predictors beyond those found with analyses of variance; four were credible, 28 feasible, and five implausible.

In both analytic frameworks, small samples truncated the impression-formation model that was uncovered in larger analyses, and thereby failed to offer evidence of certain impression-formation processes that may actually be operative. Conjunctive stepwise results from the 128-actions samples missed 52 percent of the predictors found with the 515-actions samples. Model specifications inferred via analyses of variance from the 128-actions subsamples also missed about half of the significant effects found with the 515-actions datasets. At the same time, small sample analyses with both analytic frameworks, but especially step-wise regressions, introduced terms that were completely absent in equations based on the large samples. Such "false positives" might induce a researcher to engage in unwarranted interpretation and theorizing.

A 192-actions subsample performed only slightly better than the 128-actions subsample while amounting to a fifty-percent increase in sample size, so adding a few additional vignettes in an impression formation study may not be worth the extra cost. Ultimately, only studies with large numbers of vignettes can uncover all of the subtleties of impression-formation processes.

Three issues in the impression-formation data contributed to problems in stepwise regression. First, it was impossible to achieve the ideal factorial design in which actors, behaviors, and objects had equal numbers of positive and negative *epa* values because few interpersonal behaviors are impotent or quiescent, and none are very much so. As a consequence, the variances of behavior potency and of behavior activity were lower than the variances of other predictors, reducing the statistical effectiveness of these two predictors, and of the interaction terms to which they contributed. Second, lack of a perfectly balanced design in the sample of action sentences led to some modest correlations among predictors, and these correlations ballooned when the predictors were multiplied together to create third-order interaction variables. As a result, multicollinearity was a problem, especially in analyzing the impact of third-order interactions. Third, some vignettes presented actions that were far more affectively intense than other actions, and these cases exerted excessive influence on results, sometimes generating predictor terms on their own.

The analysis of variance framework moderated all three of these problems. With events categorized into equal size categories, every factor had the same variance, 0.25. The factors still correlated because the attained design was not perfectly balanced, but, for example, dichotomization reduced the correlation in the male 515-actions data between B_e and B_a from -0.37 to -0.13, and yielded a maximum correlation among all factors of just 0.18 (B_p with B_a). Coding into dichotomies completely removed the third problem encountered with stepwise regressions. No event had more influence on results than other events because the profile of every event now was defined solely in terms of just two values. Consequently, the analyses of variance provided more parsimonious and plausible specifications of causal structure than did stepwise regressions. Additionally the analysis of variance approach accords with the argument that vignette studies of normative processes are experiments. The vignettes are generated to represent a variety of factors that might control the responses of individuals encountering the vignettes, and analysis of variance is a standard means of determining which factors and which interactions of factors contribute to response variance in an experiment.

With the analysis of variance framework, equation estimation in order to quantify effects occurs after specification of causes. That is, the analyses of variance define a structural equation model, which can be estimated with the same data as used in the analysis of variance, now in its original form of scaled measurements rather than dichotomies. Software for structural equation modeling fosters testing for differences in effects across different groups, thereby providing a practical way to determine if there are cultural or sub-cultural differences in normative process beyond those discovered through analyses of variance. Specifications of normative process in each group are pooled in order to define a model that can be applied to all groups simultaneously. Thereupon one can use group comparison capabilities of structural equation modeling in order to test for cultural differences in coefficients. For example analyses here suggested significant gender differences in sizes of some effects while analyses of variance indicated only that the effects were operative for both sexes.

Finally, it is worth mentioning that the method forwarded here—analyzing the same data first with analyses of variance and then again with structural equation modeling—sometimes might be applicable in research areas other than the study of normative processes, when exogenous variables can be viewed as factors, and the sample of cases is distributed over the factors and their relevant interactions.

Endnotes

1. Canadian data on impression formation are available at the affect control theory website (http://www.indiana.edu/~socpsy/ACT/data.html) with the specific URL of http://www.indiana.edu/%7Esocpsy/public_files/Canada_Ratings_of_214_Events.xls.

2 "Significance" lacks its usual probabilistic meaning in stepwise regression analyses because significance tests are applied serially in analyses that are structured by previous analyses of the same dataset. Thus, significance in stepwise analyses means that a standard test yields a result that would be judged as improbably null were the test's assumptions fulfilled, which they are not.

3 The 515-actions dataset is available at the affect control theory website (http://www.indiana.edu/~socpsy/ACT/data.html) with the specific URL of http://www.indiana.edu/~socpsy/public_files/UNC78_Ratings_of_515_events.xls

4 See the appendix to Heise's (2010) Chapter 4 for more details regarding the design principles.

References

Cook, R. Dennis, and Sanford Weisberg. 1999. *Applied Regression Including Computing and Graphics.* New York: John Wiley & Sons.

Efron, Bradley, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. 2004. "Least Angle Regression." *Annals Of Statistics* 32:407-499.

Gollob, Harry F. 1968. "Impression formation and word combination in sentences." *Journal of Personality and Social Psychology* 10:341-53.

Gollob, Harry F. and B. B. Rossman. 1973. "Judgments of an actor's 'Power and ability to influence others'." *Journal of Experimental Social Psychology* 9:391-406.

Jasso, Guillemina. 2006. "Factorial survey methods for studying beliefs and judgments." *Sociological Methods & Research* 34:334-423.

Heider, Fritz. 1958. *The psychology of interpersonal relations*. New York: John Wiley & Sons.

Heise, David R. 1969. "Some methodological issues in semantic differential research." *Psychological Bulletin* 72:406-422.

—. 2007. *Expressive Order: Confirming Sentiments in Social Actions*. New York: Springer.

—. 2010. *Surveying Cultures: Discovering Shared Conceptions and Sentiments*. Hoboken, NJ: Wiley Interscience.

Hoey, Jesse, Tobias Schröder, and Areej Alhothali. 2013. "Bayesian Affect Control Theory." Pp. 166-172 in *Humaine Association Conference on Affective Computing and Intelligent Interaction*. Geneva, Switzerland: Institute of Electrical and Electronics Engineers, Computer Society.

MacKinnon, Neil J. 1985. "Affective dynamics and role analysis (Report for SSHRC Project 410-81-0089)." Department of Sociology and Anthropology, University of Guelph, Guelph, Ontario.

—. 1994. *Symbolic Interactionism as Affect Control*. Albany, NY: State University of New York Press.

Osgood, Charles E., W. H. May, and M. S. Miron. 1975. *Cross-Cultural Universals of Affective Meaning*. Urbana: University of Illinois Press.

Robinson, Dawn T. and Lynn Smith-Lovin. 2006. "Affect Control Theory." Pp. 137-164 in *Contemporary Social Psychological Theories*, edited by P. Burke. Stanford, CA: Stanford University Press.

Romney, A. Kimball. 1994. "When does consensus indicate cultural knowledge?" *CogSci News* 7:3-7.

Romney, A. Kimball, William H. Batchelder, and S. C Weller. 1987. "Recent applications of cultural consensus theory." *American Behavioral Science* 31:163-177.

Romney, A. Kimball, S. C Weller, and William H. Batchelder. 1986. "Culture as consensus: A theory of culture and informant accuracy." *American Anthropologist* 88:313-338.

Rossi, Alice S. and Peter H. Rossi. 1990. *Of Human Bonding: Parent-Child Relations Across the Life Course*. New York: Aldine de Gruyter.

Rossi, Peter H. and Andy B. Anderson. 1982. "The factorial survey approach: An introduction." Pp. 15-67 in *Measuring Social Judgments: The Factorial Survey Approach*, edited by P. H. Rossi and S. L. Nock. Beverly Hills: Sage Publications.

Rossi, Peter H. and Steven L. Nock. 1982. "Measuring Social Judgments: The Factorial Survey Approach." Beverly Hills: Sage Publications.

Schröder, Tobias. 2011. "A Model of Language-Based Impression Formation and Attribution Among Germans." *Journal of Language and Social Psychology* 30:82-102.

Smith-Lovin, Lynn. 1987. "Impressions from events." *Journal of Mathematical Sociology* 13:35-70.

Smith, Herman W., Takanori Matsuno, and Michio Umino. 1994. "How similar are impression-formation processes among Japanese and Americans?" *Social Psychology Quarterly* 57:124-139.

Wallander, Lisa. 2009. "25 years of factorial surveys in sociology: A review." Social Science Research 38:505-52.