Advances in Data Analysis, Data Handling and Business Intelligence

Proceedings of the 32nd Annual Conference of the Gesellschaft für Klassifikation e.V., Joint Conference with the British Classification Society (BCS) and the Dutch/Flemish Classification Society (VOC), Helmut-Schmidt-University, Hamburg, July 16-18, 2008

Bearbeitet von
Andreas Fink, Berthold Lausen, Wilfried Seidel, Alfred Ultsch

ISBN 978 3 642 01043 9
Format (B x L): 15,5 x 23,5 cm
Gewicht: 1235 g

Wirtschaft > Betriebswirtschaft: Theorie & Allgemeines > Wirtschaftsinformatik, SAP, IT-Management

schnell und portofrei erhältlich bei

beck-shop.de
DIE FACHBUCHHANDLUNG
Strategies of Model Construction for the Analysis of Judgment Data

Sabine Krolak-Schwerdt

Abstract This paper is concerned with the types of models researchers use to analyze empirical data in the domain of social judgments and decisions. Models for the analysis of judgment data may be divided into two classes depending on the criteria they optimize: Optimizing an internal (mathematical) criterion function with the aim to minimize the discrepancy of values predicted by the model from obtained data or incorporating a substantive underlying theory into the model where model parameters are not only formally defined, but represent specified components of judgments. Results from applying models from both classes to empirical data exhibit considerable differences between the models in construct validity, but not in empirical validity. It may be concluded that any model for the analysis of judgment data implies the selection of a formal theory about judgments. Hence, optimizing a mathematical criterion function does not induce a non-theoretical rationale or neutral tool. As a consequence, models satisfying construct validity seem superior in the domain of judgments and decisions.

Keywords Model comparison · Models of data analysis · Social judgments · Validity.

1 Introduction

This paper is concerned with the types of models researchers develop and use to analyze empirical data in the domain of social judgments and decisions.

Social judgments play a central role in the professional as well as private everyday life. Social judgments are a key prerequisite for coordinated social life, and the ability to integrate complex social information for judgment purposes is one of the most demanding tasks (Fiske & Taylor, 2008). Frequently, these judgments
contribute to far reaching decisions about persons. Examples are medical expert judgments, court decisions or decisions about job applicants.

The combination of characteristics or attributes of the object to be judged into a composite, which represents the judgment, is a pervasive and important problem in nearly all kind of decision making situations (Einhorn & Hogarth, 1975). For example, it may be considered how symptoms are to be weighted and combined into a clinical judgment about the disease of a person. These problems subsume the following main aspects of the combination issue: (1) specifying the function that relates the attributes to the composite and (2) determining the weight of each attribute to the composite.

A great number of investigations have approached the measurement of judgments and possible biases or flaws in judgments. Typically, these are faced with the following problem. On the one hand, there are substantive cognitive theories and empirical results on the nature of judgments. On the other hand, the choice of a method to analyze judgment data implies the selection of a formal theory about judgments. Hence, the method has the same function of model building as the cognitive theory. Both have to coincide, otherwise artefacts are obtained instead of valid results. The term “method” refers to a statistical method to analyze the data or a model of data analysis. The question may be raised how an adequate method may be constructed.

2 Strategies of Model Construction

Corresponding methods may be divided at least into two classes depending on the criteria they optimize (Apostel, 1961; Roskam, 1979). The first is optimizing a mathematical criterion function. The aim is to minimize the discrepancy of values predicted by the model from obtained data. Frequently, least squares procedures are used. An example is multiple regression with the model equation \( y_i = \alpha + \sum_j \beta_j x_{ij} + \epsilon \). The interesting model parameters are the regression weights \( \beta_j \) which are estimated to fulfil the least squares criterion \( \sum_i (y_i - \alpha - \sum_j \hat{\beta}_j x_{ij})^2 := min \). The criterion which has to be satisfied by a valid model of the obtained data is empirical validity. Thus, empirical validity involves that the model should fit the obtained data and it is usually assessed by an overall goodness-of-fit measure. A measure that is frequently used is the correlation \( R \) between the obtained data and the predicted data or its square \( R^2 \) which specifies the amount of variation in the data predicted by the model (cf. Harshman, 1984).

In the second approach, the criterion which has to be satisfied is construct validity. Construct validity refers to the ability of the model parameters to reflect the judgment structures that they are specified to represent. The corresponding type of model construction involves that a substantive theory is incorporated into the model. An example from psychophysics concerns the modelling of taste impressions. According to the psychophysical law \( k = C^n t \), a specific taste \( k \) depends on the concentration \( C \) of a tasted liquid and duration \( t \) of exposition. The exponent \( n \) is different for tastes like sweet, bitter or sour. Thus, the model parameter \( n \) is not just
formal, but represents a specific taste. In other words, it has construct validity. Item response theory from the domain of psychological assessment is another instance, where model parameters reflect characteristics of items and of persons responding to the items of a test (Fischer, 1996).

The basic thesis of this paper is that the second approach involving construct validity should be advanced within the domain of judgments and decisions. Due to their construction, models satisfying construct validity have the potential to integrate substantive theories on the nature of judgments and methods of data analysis. In order to substantiate the superiority of the second approach, the following exposition introduces selected theories and findings on social judgments from two broad research lines first. Subsequently, it is outlined how these relate to model parameters within the two model classes introduced above by use of empirical data. Finally, implications of the findings on principles of model construction are discussed.

3 Theories of Judgment and Empirical Findings

The first research line concerns the way people integrate pieces of information for judgment purposes. Judgments consist of gradations along a number of dimensions such as valence or agreeableness of persons (e.g., judgments about the degree a person is friendly or unfriendly, idealistic or materialistic, talented or dull) (Anderson & Sedikides, 1991). In a number of judgment conditions people make judgments based on all of the relevant information, weighted and combined into a dimension by an algebraic integration principle (Anderson, 1981; Fiske & Taylor, 2008). This principle has been stated for the first time by Benjamin Franklin (cited from Dawes & Corrigan, 1974, p. 95):

My way is to divide half a sheet of paper by a line into two columns; writing over the one Pro, and over the other Con. Then, ... I put down under the different heads short hints ... for or against the measure. When I have thus got them all together in one view, I endeavor to estimate the respective weights ... to find at length where the balance lies.

The principle may be stated formally as \( y_i = \sum b_j x_{ij} \) and is nowadays known as Franklin’s rule (cf. Gigerenzer & Todd, 1999). Within this basic rationale, empirical results have shown that people use weights \( b_j \) of \(+1\) and \(-1\) to form the judgment, termed the unit weighting principle (Bröder, 2002; Dana & Dawes, 2004). Thus, people simply add information with positive evidence for the judgment (i.e., Pros) and subtract information with negative evidence (i.e., Cons).

The second research line is concerned with the effects that existing knowledge structures in memory have on judgments. In the social domain, such knowledge structures comprise categories or stereotypes. Stereotypes are cognitive structures that contain peoples’ knowledge, beliefs and expectations about social groups (e.g., Fiske & Taylor, 2008). They involve illusory correlations of category membership and specific attribute domains (Hamilton & Gifford, 1976). Thus, they create connections between judgment dimensions which are statistically independent. As an
example, in the stereotype of a “skinhead” agreeableness and dominance are negatively correlated. Thus, a person categorized as a skinhead is judged as not agreeable and dominant. Hence, stereotypes bias judgments by causing such correlations.

The following exposition outlines both research domains in detail before returning to the question of model construction.

3.1 The Problem of Information Weighting

As to the question how information is weighted and combined into a judgment, behavioral decision research is confronted with the problem of drawing conclusions about unobservable decision strategies from behavioral data. Strategies like the unit weighting principle or the standard multiple regression model are competing theories about information integration in judgment and decision tasks. The design of studies which have the aim to draw corresponding conclusions consists of regressing judgment data as criterion values on the presented pieces of information as a set of predictors by either strategy and subsequently comparing the variances accounted for.

Empirical evidence which shows that the unit weighting principle is superior to the regression model with optimal regression weights in approximating human judgments comes from experimental and field studies within functional measurement theory (Anderson, 1981), social judgment theory (Stewart, 1988), simulation studies (Bröder, 2002) and some other domains. In these studies, unit weighting models are superior in the sense that they are more parsimonious than regression with optimal weights, but have comparable empirical validity. Consequently, in order to formulate an adequate, that is, a frugal model for the analysis of judgment data, the weights may be restricted to unity without much loss of information.

Even more intriguing is the fact that unit weighting models correlate highly and in a number of studies nearly perfectly with the predictions from standard regression analysis. Stated in other words, it has been repeatedly demonstrated that the unit weighting strategy is fairly accurate as compared to regression models with optimal regression weights (Bröder, 2002; Dawes & Corrigan, 1974; Wainer, 1976).

Finally, and most importantly, unit weighting yields a valid prediction of a known, true criterion. Far-reaching findings were presented by Dawes and Corrigan (1974). In a number of studies, they have used large empirical data sets from clinical psychology, education and perception. In the following, their procedures and results will be very briefly outlined.

In Study 1, first-year graduate students in the department for psychology at the University of Illinois were evaluated on 10 variables which were predictive of academic success. These variables included aptitude test scores, college grade point average, peer-ratings on extroversion and self-ratings on conscientiousness. A Graduate record exam (GPA) was computed for all these students. This served as the external validity criterion. The aim of the study was to predict the GPA results from the 10 variables.
In Study 2, graduate students in the department of psychology at the University of Oregon, who had been there for at least two years, were evaluated on a five-point-rating scale by faculty members who knew them well. This was one external validity criterion. The other criterion was the final Graduate record Exam (GPA).

At the time the students applied, three scores were obtained for each student: His or her Undergraduate record Exam (GRE), undergraduate grade point average and a score of the quality of the institution at which the undergraduate exam has been passed. These scores were available to the admission committee at the time the students applied and they served as predictors. The problem was twofold: (1) To predict the final Graduate record exam (GPA) from these three variables and (2) to predict the ratings of the faculty members from these variables.

In Study 3, which was an experiment on perception, participants received ellipses which were varied on the basis of each figure’s size $i$, eccentricity $j$, and grayness $k$. The formula for variation used by the experimenters was $ij + jk + ik$. Participants’ task was to estimate the value of each ellipse. The external validity criterion was the true (that is, experimenter assigned) value of each ellipse on the basis of its size, eccentricity, and grayness.

In all of these studies, the data analysis was the following. The problem was always to predict the external validity criterion. Predictors were integrated for the prediction by several models. One was by use of estimating optimal beta weights in a standard regression analysis and another one was by use of a unit weighting to integrate the predictors. Table 1 shows the validity results from applying both models.

Results from the experimental Study 3 show identical validity coefficients for both models (see Table 1). In the other studies, validity of the optimal linear model is increased as compared to the unit weighting scheme, but only slightly increased. Thus, the difference in validity coefficients between the two models does not really matter. The conclusion drawn from this and many other studies is that researchers in the domain of judgments should not bother about an optimal model at all (Bröder, 2002; Dawes & Corrigan, 1974; Wainer, 1976). Obviously, unit weights are as predictive as regression procedures (e.g., Einhorn & Hogarth, 1975; Schmidt, 1972; Claudy, 1972).

The obvious next question is then: What are the reasons for the unit weighting model to be an adequate approximation for human decision behavior and to

<table>
<thead>
<tr>
<th>Study</th>
<th>Validity of unit weighting model</th>
<th>Validity of optimal linear model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Illinois students’ predictions of GPA</td>
<td>0.60</td>
<td>0.69</td>
</tr>
<tr>
<td>2: Oregon students’ predictions of GPA</td>
<td>0.69</td>
<td>0.69</td>
</tr>
<tr>
<td>2’: Oregon faculty members’ ratings</td>
<td>0.48</td>
<td>0.54</td>
</tr>
<tr>
<td>3: Predictions of ellipses</td>
<td>0.97</td>
<td>0.97</td>
</tr>
</tbody>
</table>
be externally valid? At least three reasons have been discussed (see Einhorn & Hogarth, 1975 for a review): (1) The weighting problem is subsidiary to specifying the relevant variables which should be put into the model. That is, once the relevant variables are included in the model, their weighting may not be very important (cf. Dawes & Corrigan, 1974). (2) In estimating optimal regression weights, there will inevitably be sampling error. In contrast, unit weights have no sampling error. Hence, there may be a potential trade-off between accuracy of estimation and estimation without error (or nearly without error). As Einhorn and Hogarth (1975, p. 173) put it, “because judgment data will contain both sampling and measurement error, the relative superiority of regression procedures over unit weighting may be quite small (or nonexistent)”.

Thus, unit weighting models are adequate in the sense that they reflect components of human judgment processes in a more parsimonious way and in the sense that the prediction is as externally valid as those of standard regression. As opposed to Franklin’s rule, which represents the standard regression model, unit weighting is nowadays well known as Dawes’ rule (Gigerenzer & Todd, 1999).

3.2 Illusory Correlations in Judgments

Dawes’ rule turns out as valid in conditions, where people make judgments based on all of the relevant information which is then integrated attribute by attribute. This holds true for all conditions where judgments have significant consequences and thus make people accountable for their decisions or when people have enough processing capacity at their disposal to revisit all the given information (Fiske & Taylor, 2008).

Under other conditions, however, people cannot afford the large processing abilities which are assumed by fully integrating the information pieces or they are not motivated to do so. In these cases, people use heuristic strategies, which are much simpler, but still represent viable alternatives (Gigerenzer & Todd, 1999). There are a number of heuristics which cannot be discussed in the present context due to space limitations. However, one prevalent strategy is to base one’s judgments on stereotypes which involve illusory correlations of attribute domains which are indeed statistically independent (Hamilton & Gifford, 1976).

The cognitive basis is that people overestimate the frequency of co-occurrence of events which are statistically infrequent: If one group of persons occurs less frequently than another and one type of behaviors occurs infrequently, then observers overestimate the frequency that this type of behavior was performed by members of that group.

In an experiment, Hamilton and Gifford (1976) presented statements about members of two groups, which were simply labeled Group A and Group B. The stimulus set contained twice as many statements about Group A as about Group B. Furthermore, the behavioral statements were desirable or undesirable, with twice as many desirable as undesirable behaviors in each group. Because the proportions of
Table 2  Experiment of Hamilton and Gifford (1976)

<table>
<thead>
<tr>
<th></th>
<th>Frequency of stimulus sentences:</th>
<th>Frequency estimates means:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group A</td>
<td>B</td>
</tr>
<tr>
<td>Behaviors:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Desirable</td>
<td>18</td>
<td>9</td>
</tr>
<tr>
<td>Undesirable</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>∑</td>
<td>26</td>
<td>13</td>
</tr>
</tbody>
</table>

desirable and undesirable statements were identical for the two groups, there is no correlation between group membership and desirability in the stimulus set. Participants had to estimate the number of undesirable behaviors in each group. Table 2 shows the frequency of the statements and the estimates of the participants. The results show an overproportionally high estimate of undesirable behaviors for Group B as compared to Group A. Phi coefficients between group membership and desirability were significant which indicates an erroneous perception of an association between the smaller Group B and undesirable behaviors (Hamilton & Gifford, 1976).

Having revisited two important research domains on social judgments, the following exposition focuses on the question, if there are models of data analysis which may incorporate this way of theorizing.

If construct validity is the criterion to be optimized in the construction of a model of data analysis, then an adequate formal approach must incorporate the following components: (1) A formal model must include the unit weighting principle to integrate person information. (2) The model has to specify parameters for the correlations of judgment dimensions in different conditions.

An appropriate model class might be three-way two-mode multidimensional scaling. These models have the potential to reflect the occurrence of correlated or independent judgment dimensions due to stereotype use conditions.

4  Three-Way Two-Mode Models

Three-way two-mode scaling models may be subdivided into the two model classes that were introduced at the outset. Thus, there is one class optimizing an internal (mathematical) criterion function. An example is the Tucker model (Tucker, 1972). In contrast, a model which was derived to satisfy construct validity is the SUMM-ID approach (Krolak-Schwerdt, 2005). What the models have in common, are the input data and the basic model equation for the data. In the following, the scalar product form of the models will be outlined.
The input data consist of a three-way data matrix $X = (x_{ijj'})$, $i = 1, \ldots, I$, $j, j' = 1, \ldots, J$, where $I$ is the number of individuals or conditions and $J$ the number of attributes. $X$ can be thought of as comprising a set of $I(\geq 2)J \times J$ scalar products matrices. $X_i$, a slice of the three-way matrix, consists of scalar products between attributes $j, j'$ for an individual or a condition $i$.

The basic model equation can be expressed as $X_i = BH_iB' + E_i$, where $B$ is a $J \times P$ matrix specifying an attribute space or judgment configuration which is common to all individuals or conditions where $P$ is the number of dimensions. $H_i$ is a $P \times P$ symmetric matrix designating the nature of individual $i$’s representation of the judgment dimensions. Diagonal elements $h_{ipp}$ of $H_i$ correspond to weights applied to the judgment dimensions by individual $i$, while off-diagonal elements $h_{ipp'}$ are related to perceived relationships among the judgment dimensions $p$ and $p'$. Matrix $H_i$, termed core matrix (Tucker, 1972), transforms the common judgment space into the individual representation, and $E_i$ collects the errors of approximation $e_{ijj'}$.

Thus, the basic model equation assumes that there is a common space represented by matrix $B$ which underlies judgments in general. On the basis of the common space, the model allows for two kinds of distortions in individual representations. The first is that individuals may attach different weights to different judgment dimensions. More important in the present context is the second type of distortion: Individual representations may be rotated versions of the common space in which independent dimensions become correlated.

To return to the two model classes, there are a number of differences between them, but in the present context the most important may be sketched in the following way: Models optimizing an internal criterion function such as the Tucker model determine the parameter matrices such that the discrepancy between obtained and predicted data will be minimized, $\sum_i \sum_j \sum_{j'} (x_{ijj'} - b_{j'p}h_{ipp'})^2 := min$. This is accomplished by a principle component analysis in the attributes’ mode or by an alternating least squares approach. The important fact for the present research question is that any real valued estimates for the entries of $B$ are considered as long as the discrepancy function is minimum. In terms of the judgment process this implies Franklin’s rule.

In contrast, SUMM-ID integrates the unit weighting principle. To sketch the underlying rationale very briefly, sign vectors $z_p$ for the attributes $j, z_{jp} \in \{-1, 1\}$, and, in an analogous way, sign vectors $s_i$ for the individuals or conditions $i, s_{ip} \in \{-1, 1\}$ are introduced, where $\sum_i \sum_j s_{ip}z_{jp}z_{j'p}x_{ijj'} = t_p := max$. An estimate of $B = (b_{j'p})$ is obtained by $b_{j'p} = \sum_i \sum_{j'} s_{ip}z_{jp}x_{ijj'}t_p^{2/3}$. For a more thorough discussion of the model, the reader may be referred to Krolak-Schwerdt (2005).

Results from applying both approaches, the Tucker model and SUMM-ID, may be fundamentally different as to construct validity of the model parameters. This will be demonstrated in the following by applying both approaches to experimental data.
4.1 Application

In an experiment on the topic of teachers’ achievement judgments of students’ performance, experienced school teachers as participants received case reports about students as materials. Each case report contained information on social activity, discipline, capability and motivation of the student. Teachers were told to form an impression of the student as in usual classroom assessments. The experiment had a factorial design where the first factor was activation of a stereotype. That is, in one of two experimental conditions a stereotype (e.g., “the student is a bloomer”) was activated prior to the presentation of the case report, while the other condition proceeded without stereotype activation (termed “non-stereotype” in the following). The second factor was replication with two other descriptions presented with or without stereotype activation.

After having read one of the descriptions, subjects had to rate the case report on seven rating scales such as capability of achievement, work ethics and so on. These scales correspond to the following dimensions: (1) social competence, (2) reasoning, (3) language capability.

For each case report, the normalized distance matrix between the scales was used as input for the data analysis. These data were then subjected to SUMM-ID and the Tucker approach. We expected the following results: A common judgment space should occur which consists of the three a-priori dimensions just mentioned. Illusory correlations should appear in increased off-diagonal values of the core matrix in the two stereotype conditions. In the other conditions, these values should be near zero.

4.1.1 Results

As to explained variation in the data, both models showed an excellent recovery of the data. The Tucker approach with 94% is slightly superior to SUMM-ID with 92%. In the SUMM-ID solution, we found three dimensions in the common judgment space. After Varimax rotation, these reflected the expected dimensions (that is, language capability, social competence and reasoning).

The core matrix1 is shown in Table 3. The off-diagonal core values exhibit the expected pattern: Judging in the presence of a stereotype yields illusory correlations in both description sets. Thus, by use of the stereotype “bloomer” teachers attribute high language and reasoning capabilities coupled with high social competence. Also in accordance with our hypothesis, we find rather independent dimensions in the non-stereotype conditions. That is, different judgment domains are used in a more unconfounded manner.

From the Tucker approach, the expected judgment dimensions were also obtained. However, the off-diagonal values of the Tucker core matrix which are shown in Table 4 do not show a systematic pattern of high vs. low correlations due

---

1 Diagonal values of the core matrix will not be discussed in the following, as they do not contribute to the estimation of the models' construct validity in the present context.
Table 3  Core matrix of the SUMM-ID solution for the school achievement data

<table>
<thead>
<tr>
<th>Experimental conditions</th>
<th>Judgment dimensions</th>
<th>Language capability</th>
<th>Social competence</th>
<th>Reasoning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Language capability</td>
<td>0.72</td>
<td>0.12</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>Social competence</td>
<td>0.68</td>
<td>0.03</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>Reasoning</td>
<td>0.05</td>
<td>0.03</td>
<td>0.66</td>
</tr>
<tr>
<td>Set 1: Non-stereotype</td>
<td>Language capability</td>
<td>0.61</td>
<td>0.74</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>Social competence</td>
<td>0.52</td>
<td>0.55</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>Reasoning</td>
<td>0.05</td>
<td>0.03</td>
<td>0.66</td>
</tr>
<tr>
<td>Set 1: Stereotype</td>
<td>Language capability</td>
<td>0.77</td>
<td>0.74</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Social competence</td>
<td>0.57</td>
<td>0.47</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>Reasoning</td>
<td>0.34</td>
<td>0.55</td>
<td>0.56</td>
</tr>
<tr>
<td>Set 2: Non-stereotype</td>
<td>Language capability</td>
<td>0.79</td>
<td>0.14</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Social competence</td>
<td>0.77</td>
<td>0.77</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>Reasoning</td>
<td>0.01</td>
<td>0.07</td>
<td>0.95</td>
</tr>
<tr>
<td>Set 2: Stereotype</td>
<td>Language capability</td>
<td>1.46</td>
<td>0.74</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>Social competence</td>
<td>1.00</td>
<td>0.48</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td>Reasoning</td>
<td>1.26</td>
<td>0.03</td>
<td>1.26</td>
</tr>
</tbody>
</table>

Table 4  Core matrix of the Tucker model for the school achievement data

<table>
<thead>
<tr>
<th>Experimental conditions</th>
<th>Judgment dimensions</th>
<th>Language capability</th>
<th>Social competence</th>
<th>Reasoning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Language capability</td>
<td>1.90</td>
<td>0.33</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Social competence</td>
<td>1.20</td>
<td>0.16</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Reasoning</td>
<td>1.01</td>
<td>0.16</td>
<td>1.09</td>
</tr>
</tbody>
</table>

To stereotype use. Rather, some high values are found in the non-stereotype condition and some low ones in the stereotype condition. Thus, the values do not indicate an increase in the magnitude of correlations between dimensions due to stereotype activation.
In conclusion, there is no correspondence between the parameters obtained from the Tucker model and the experimental manipulations. In contrast, the SUMM-ID model reflects the expected structure of the common judgment space and the expected distortions of this space due to stereotype activation in every detail. That is, the parameters of the approach were sensitive to manipulations of stereotype activation and thus have construct validity.

5 Conclusions

At the outset, we distinguished two ways of model construction: Optimizing a mathematical criterion function which is the usual approach or integrating an underlying theory such that the model parameters have construct validity. As to empirical validity, results from both approaches were comparable.

As to construct validity, the second approach turns out to be superior. The reason is that optimizing a mathematical criterion function does not induce a non-theoretical rationale. Rather, this approach yields another formal theory about judgments which does not correspond to substantive theories. People simply do not consider all possible weights for person information, but only a very limited number. Thus, the unit weight rule is a better predictor of people’s judgments than Franklin’s rule. In more general terms, optimizing construct validity guarantees a close correspondence of the formal model to the underlying substantive theory. As a consequence, the model extracts the theoretically significant parts from the data. As the presented empirical results have also shown, this does not imply to give up optimal predictions in the sense of minimizing discrepancies of predicted from obtained data. In conclusion, then, the second strategy of model construction should be focused more in future research if the aim is to develop valid models of the corresponding research domain.

References


