

INCOMPATIBILITY IN THE RELATIONSHIPS BETWEEN COVARIATION INFORMATION AND ATTRIBUTIONAL DIMENSIONS

Frank VAN OVERWALLE
Vrije Universiteit Brussel

This article examines four distinct relationships between covariation patterns and attributional dimensions (*consensus* -> *locus*; *consistency* -> *stability*; *distinctiveness* -> *globality*; *action-contingency* -> *control*) as specified in the Joint Model by Van Overwalle and Heylighen (1995), using an adapted stimulus-response compatibility paradigm. Subjects were first told that a person was rejected after an application interview and then given additional covariation information. Next, they had to judge whether a particular cause was the "correct" explanation for this rejection, on the basis of a causal rule given before. This rule was either compatible or incompatible with the joint model's predictions. Assuming that this model accurately captures subjects' natural causal judgments, it was predicted that incompatible causal rules should be more difficult, leading to more errors against the imposed rule and longer response times, whereas rules that are compatible should be easier, leading to less errors and shorter response times. Two experiments confirmed the predicted covariation-attribution relationships, except for the contingency-control relationship which was confirmed only by the accuracy data.

The ability of humans to identify the causes of various events is indispensable to control the environment, to successfully cope with difficulties and to predict the future. According to Kelley (1967), causal inferences are determined by the principle of covariation which says that "the effect is attributed to that condition which is present when the effect is present and which is absent when the effect is absent" (p. 194). In later theorizing and research, this principle has been refined and extended in several covariation theories (e.g., Cheng & Novick, 1990; Hilton & Slugoski, 1986; Hewstone & Jaspars, 1978; Försterling, 1989).

Although most data generally supported Kelley's original conception and later developments, Van Overwalle and Heylighen (1995) claimed that this earlier research and theory is limited in that it often restricted people's causal thinking to three potential attribution categories. Specifically, people were assumed to make attributions almost exclusively to the *person*, the *occasion*, the *stimulus* or a combination of these three factors. This stands in sharp contrast with findings from research on the dimensionality of spontaneous

I am grateful to Pieter Janssen and Sarah Casaer for serving diligently as experimenters, and to Dirk Van Rooy for his comments on an earlier version of the manuscript.

Correspondence concerning this article should be addressed to Frank Van Overwalle, Vrije Universiteit Brussel, Department of Psychology, Pleinlaan 2, 1050 Brussels, Belgium. Electronic mail may be sent to Frank.VanOverwalle@vub.ac.be.

every-day attributions, which revealed that causes can be categorized along at least four dimensions, including *locus* (internal or external to the agent), *stability* (stable or variable over time), *globality* (general or situation-specific) and *control* (changeable or unchangeable by one's actions; see Abramson, Seligman & Teasdale, 1978; de Jong, Koomen, & Mellenbergh, 1988; Van Overwalle, 1989; Weiner, 1986). Theory and research in applied fields also support this four-dimensional structure for explaining the psychological consequences of attributions such as motivation and emotion (Weiner, 1986), learned helplessness (Abramson et al., 1978; Peterson & Seligman, 1984), consumer behavior (Folkes, 1984), and marriage (Bradbury and Fincham, 1990). Although the four proposed dimensions can be viewed as continuous variables ranging from one extreme to the other, most often they are simply dichotomized to reflect eight possible causal categories. Clearly, three of these categories are also represented in prior covariation theories, as attributions to the "person" are identical to internal causes, attributions to the "occasion" are similar to variable causes, and attributions to the "stimulus" are identical to specific causes. However, the opposite dimensional categories are not represented, and the notion of control is completely absent.

To include these additional attributional categories in a more comprehensive model of the attribution process, Van Overwalle and Heylighen (1995) proposed that people make attributions to all eight attributional categories by applying J. S. Mill's (1872/1973) *method of difference* and *method of agreement*. Since Mill defined the combination of these two methods as the joint method, Van Overwalle and Heylighen (1995) termed their model likewise the *joint model*. Their emphasis on Mill's two methods differs from earlier covariation models which are basically derived from the method of difference only (e.g., Hilton & Slugoski, 1986; Hewstone & Jaspars, 1978; Försterling, 1989; Cheng & Novick, 1990, 1991) and left relatively unspecified the complementary method of agreement (but see Hilton, Smith & Kim, 1995; Van Overwalle, 1997b).

The method of difference compares an event where the effect is present to other events in which the same effect is absent, and specifies the cause as the one condition that differs between these two events.¹ Imagine, for example, a girl - Mary - who applies for a job as a city guide. The information that Mary is the only candidate who was rejected suggests that Mary "differs" from other people and designates her as the cause. In one of the most influential covariation models in social psychology, Cheng and Novick (1990) proposed the concept of a probabilistic contrast as an instantiation of the method of difference which

¹ It would be more correct to denote this Mill's *indirect* method of difference, because the target event is compared to a *set* of comparison events rather than just one *single* comparison event. However, since this distinction is of less relevance here, for the sake of simplicity we will use the term *method of difference* to indicate also the indirect method.

compares events where the candidate cause is present versus absent, and infers its causal strength from the difference between the probabilities of the effect in these two cases.

In attribution research, information denoting a difference or contrast has typically been provided using Kelley's (1967) three dimensions of covariation information: Low consensus indicates that the person has a different outcome than other persons (e.g., Mary was the only person who was rejected), low consistency indicates that the outcome differs on this occasion (e.g., only applicants who applied today were rejected) and high distinctiveness reflects that the outcome is different for this stimulus (e.g., Mary was rejected only for her application as a city guide, but not for other jobs). It has been repeatedly found that such contrast information generates an attribution to the contrastive person (e.g., Mary), occasion (e.g., today) or stimulus (e.g., job as city guide) respectively. Alternatively, as noted earlier, Van Overwalle and Heylighen (1995) assumed that these attributions can be conceived as reflecting *internal*, *variable* and *specific* causal categories. Since these causes are inferred when there is a difference or contrast, they denoted these as *contrast causes*.

There are, however, instances in which the outcome of a series of events does not change, for example, when all job applicants are rejected. In such cases, J. S. Mill (1872/1973) proposed the *method of agreement* to infer a causal judgment. The method of agreement compares events in which the same effect is invariantly present, and specifies the cause as the condition that is also invariantly present ("agrees") in these events (see also Hilton, Smith & Kim, 1995). According to the joint model, high consensus information indicating that all applicants were rejected suggests that some *external* circumstance (e.g., high job requirements) is responsible, because only factors external to the applicants are invariantly present with a uniform rejection decision. Similarly, high consistency indicating that applications were rejected on all days suggests a relatively *stable* explanation (e.g., there were no guides needed any more), and low distinctiveness indicating that Mary was rejected for all jobs suggests a *general* explanation (e.g., Mary lacks the required capacities). Since the method of agreement implies that these causes are inferred when there is no difference provided in the information, Van Overwalle and Heylighen (1995) defined these as *invariant causes*.

It is important to note that Mill (1872/1973) regarded the method of agreement as weaker than the method of difference, because there are in principle an almost infinite number of invariantly present conditions (e.g., oxygen, gravitation, and the like) which are *not* causally related to a given effect (e.g., rejection of all applicants). Despite this limitation, Mill proposed that the method of agreement can be used to narrow down the causal field. Thus, perceivers might conclude that the cause lies within certain attributional categories (e.g., external, stable or general). Within a given story context such

as rejection at a job interview, prior knowledge may further aid to select a more specific explanation as a most plausible candidate satisfying these attributional categories.

In research testing the method of agreement, subjects were typically requested to judge broad attributional categories, to rate specific explanations which fitted the story context, or to provide free causal explanations (which subjects presumably would select on the basis of their relevance with respect to the story; Van Overwalle & Heylighen, 1995; Van Overwalle, 1997a, 1997b). Given these response options, the results of these studies showed strong support for the application of the method of agreement. Moreover, the method of agreement is crucial for making attributions about dispositions of actors and objects, as demonstrated by Hilton, Smith and Kim (1995) and Van Overwalle (1997b).

However, the joint model's prediction for invariant causes is in contradiction with other existing covariation models which do not incorporate these causal response options nor the method of agreement. These models make either no predictions in these invariant cases (e.g., Hewstone & Jaspars, 1978; Försterling, 1989) or predict that no attributions are made *unless* additional contrast information is available (Cheng & Novick, 1990, 1991). Specifically, Cheng and Novick predict that a factor is designated as a causally irrelevant or enabling condition depending on the default assumptions people make about events outside the given or focal set of information (e.g., recruited from prior knowledge). When these default assumptions do not involve a contrast with the focal information, then the factor is irrelevant; when they do, then the factor is an enabling condition. Because it is very unlikely that everyone holds the same default assumptions for an event, these definitions imply that some people will make an irrelevant causal judgment, whereas others will make an enabling inference, resulting in weaker attributions overall given invariant information. From a different perspective, Van Overwalle and Heylighen (1995) also demonstrated that Cheng and Novick's (1990) probabilistic calculus actually predicts that any attribution given invariant information should be weaker than given contrast information (Van Overwalle & Heylighen, 1995, p. 448, footnote 5).

In addition to Kelley's (1967) covariation variables, Van Overwalle and Heylighen (1995) extended the joint model to a novel covariation variable denoted as *action-contingency*. They tentatively proposed that this variable might be diagnostic for inferring the degree of *control* over the cause (Weiner, 1986). The contingency variable was originally proposed by Abramson et al. (1978) and compares the effect across different behaviors, in analogy with Kelley's other covariation variables which compare the effect over persons, time or stimuli. Low contingency indicates that the effect remained the same whatever a person's actions (e.g., Mary was repeatedly rejected after a job

interview regardless of her efforts); whereas high contingency indicates that some action was successful in changing the effect while other actions were not (e.g., she was accepted only after she changed her clothing). According to the joint model, the contingency information reflects the probability that a certain action leads to the effect and hence determines the perceived controllability of the cause (Weiner, 1986). If contingency is low (e.g., Mary's actions were unsuccessful) the effect is seen as caused by an uncontrollable factor; conversely, if contingency is high the factor is seen as controllable. Although Overwalle and Heylighen (1995) argued that action-contingency is necessary to infer controllability, they also cautioned that contingency alone may be insufficient. Indeed, in order to infer full control, one typically also assumes that the actor had intention and foreknowledge about the effects of his or her actions (Heider, 1958). None of these notions is incorporated in the contingency variable. Hence, it is quite possible that this novel variable is less fundamental than Kelley's (1967) three original covariation variables.

To summarize, the joint model proposes a one-to-one relationship between covariation variables and causal categories, that is, each information variable is linked with one and only one causal dimension (see Table 1). As can be seen, when the pattern of information involves a contrast, by applying the method of difference, attributions are made to contrastive causes (i.e., personal factor, variable occasion, specific stimulus, controllable action). Conversely, when the pattern is invariant, by applying the method of agreement, attributions are made to invariant causes (i.e., external, stable, global and uncontrollable factors).

Table 1
Predictions of the Joint Model

Information	Variables	Factor	Pattern	Causal categories
Consensus	low	Persons	Contrast	Internal (<i>Person</i>)
	high		Invariant	External
Consistency	low	Time	Contrast	Variable (<i>Occasion</i>)
	high		Invariant	Stable
Distinctiveness	high	Stimuli	Contrast	Specific (<i>Stimulus</i>)
	low		Invariant	Global
Contingency	high	Actions	Contrast	Controllable
	low		Invariant	Uncontrollable

Note. Between parentheses on the right are the standard attribution responses predicted by earlier theory and research (e.g., Cheng & Novick, 1990)

In earlier research which confirmed these relationships (Van Overwalle & Heylighen, 1995; Van Overwalle, 1997a), subjects were given covariation information and then asked to provide a causal judgment. This is the method commonly used in previous attribution research. The aim of this article, however, is to explore these relationships using a different paradigm adapted from research on stimulus-response compatibility (see Kornblum, Hasbroucq, & Osman, 1990). Subjects were given instructions on how to relate the covariation information to the "correct" attribution response. These imposed rules were either compatible or incompatible with the covariation-attribution links predicted by the joint model. For instance, the compatible version of the rule relating consensus information with locus attributions, instructed the subjects to attribute low consensus to internal causes and high consensus to external causes, as postulated in the joint model. In contrast, in the incompatible version of the rule these relations were reversed: Low consensus had to be attributed to external causes and high consensus to internal causes. Subjects were then given potential causes that were internal or external (as ascertained by pretesting), and had to indicate as soon as possible whether each cause was "correct" or not according to the rule instructions given before.

Kornblum et al. (1990) postulated that when the task instructions are compatible with subjects' prior knowledge and experience, then the response is rapidly and accurately given. However, when the instructions are not compatible and conflict with subjects' natural responses, this will increase the difficulty of the task, so that the responses are relatively slow and error prone. If the joint model captures subjects' natural attribution processes as proposed by Van Overwalle and Heylighen (1995), then these same predictions can be extended to the present attribution task. Specifically, under compatible instructions (which agree with the joint model), I expect the attribution judgments to be relatively natural and easy, so that subjects' responses will be rather quick and accurate. In contrast, under the incompatible conditions (which conflict with the joint model), the judgments will be more artificial and difficult, so that their responses will be slower and less accurate. In line with the joint model but contrary to prior covariation models, I further predict that these compatibility differences will hold regardless of whether the covariation information implies a contrast or invariant pattern, or whether the judgment involves a contrast or invariant cause.

Experiment 1

Method

Subjects. Subjects were 56 freshmen students of the Psychology department of the Vrije Universiteit Brussel who participated as a partial requirement for

an introductory psychology course. The experiment was announced as a study on "how people make causal inferences", and lasted about half an hour.

Overview and procedure. The experiment took place in a private room, and was conducted with one or two subjects. They were seated in separate booths in front of an IBM compatible computer with a color screen, on which the instructions, the practice trials and the experimental trials appeared. Subjects were told that they would receive information about different individuals who applied for a job as a city guide. On each trial, they learned that after an interview with the personnel manager, the applicant was not engaged. Their task was to determine the cause for this rejection, using a rule which had to be followed strictly in order to make their decision. The experiment consisted of eight blocks of trials, each involving a different judgment rule given at the beginning of the block. Four rules involved the four relationships between the covariation variables and the causal dimensions as proposed in the joint model (e.g., between consensus and locus) and are termed *compatible* rules. In the four remaining rules, the relationships in each of the compatible rules were reversed so that they were *incompatible* with the joint model.

At each trial, the outcome was shown ("[the actor] was not engaged") together with the type of information described in the rule, either in a contrastive of an invariant pattern (e.g., low or high consensus). Next, a potential explanation appeared which tapped the relevant dimension and which involved either a contrast or an invariant cause (e.g., internal or external). Subjects had to respond as fast as possible whether this cause was the correct explanation according to the rule given, by pressing a yes or no key. For each block, the two levels (high and low) of the information variable were combined with two contrast causes and two invariant causes of the corresponding dimension (see Table 2), resulting in eight trials per block or 64 trials over all eight blocks. The order of the blocks and the order of the trials within each block were randomized for each subject by the computer. The name of the actor was randomly selected out of 50 common male and female Flemish first names.

Compatible - incompatible judgment rules. Each block began with a rule instructing subjects how to use the covariation information to decide about the "correct" cause. As noted earlier, four blocks specified compatible rules whereas four other blocks specified incompatible rules. The rules prescribed which type of covariation information appeared in the next block of trials and defined what type of causes should correspond with it. Each rule is denoted by the covariation information involved. For example, the compatible version of the consensus rule instructed subjects to attribute low consensus to internal causes and high consensus to external causes:

If everyone, EXCEPT this person, was engaged in this selection round, then the correct explanation is an INTERNAL cause. INTERNAL causes are factors which reside inside the applicant, which reflect an aspect of him- or herself.

Table 2
Cue Words for Causal Explanations with their Dimensional Properties

Cue Words	Causes
NO INTEREST	The candidate has no interest in a job as guide (i - - c)
NOT PREPARED	The candidate was badly prepared for the interview (i v s c)
APTITUDE JOB	The candidate lacks necessary aptitude for this job (- f - -)
SUDDENLY ILL	The candidate became suddenly ill at the interview (- v - -)
GENERAL IQ	The candidate has low general intelligence (- - g -)
PHYSICALLY HANDICAPPED	The candidate is physically handicapped (- f g u)
UNFRIENDLY MANAGER	The personnel manager is unfriendly (e - s u)
STANDARDS MANAGER	The personnel manager set high standards (e - - -)

Note. Dimensional categories are indicated between parentheses; i = internal, e = external, v = variable, f = fixed/stable, s = specific, g = global, c = controllable, u = uncontrollable. There were 4 cue words for each dimension (2 for each extreme) which showed the most clear-cut ratings (from Van Overwalle & Heylighen, 1995, exp. 4). These categories were not shown to the subjects.

If NOBODY of the applicants was engaged in this round, then the correct explanation is an EXTERNAL cause. EXTERNAL causes are factors that reside outside the applicant, which reflect an aspect of the situation or of other persons.

In the incompatible version of the consensus rule, the "correct" correspondence between consensus and locus was reversed so that low consensus had to be attributed to external causes and high consensus to internal causes:

If everyone, EXCEPT this person, was engaged in this selection round, then the correct explanation is an EXTERNAL cause. EXTERNAL causes are factors that reside outside the applicant, which reflect an aspect of the situation or of other persons.

If NOBODY of the applicants was engaged in this round, then the correct explanation is an INTERNAL cause. INTERNAL causes are factors which reside inside the applicant, which reflect an aspect of him- or herself.

The instructions for the other rules were similar. On the basis of these rules, subjects had to judge as quickly as possible whether a particular cause was correct by pressing a yes or no key. To ensure that they strictly followed these instructions, during the whole block, a card with the rule instruction was placed below the screen, and feedback on the accuracy of the response was given after each trial. To familiarize the subjects with this procedure, before the experiment, eight practice trials were given with the rule that subjects could respond to their own liking.

Covariation information. After the trial began with the outcome (i.e., rejection) displayed on the screen, and after pressing the space bare, one piece of covariation information was shown. The exact wording of the covariation information is illustrated here for each of the four possible covariation variables and their two possible invariant or contrast patterns respectively (shown in italics):

- *Nobody/Everybody except [actor] was engaged [high/low consensus]*
- *[actor] was never/always engaged after prior interviews [high/low consistency]*
- *[actor] was not/was engaged by other companies [low/high distinctiveness]*
- *[actor] was not engaged regardless what [s]he did or tried/was engaged when [s]he tackled the interview differently [low/high contingency].*

After reading the covariation information for 700 msec and the focus question "Is the cause" for another 300 msec, two cue words representing the cause appeared on the screen. Response time was recorded from that moment until the yes or no key was hit.

Causes. The causes involved eight concrete everyday explanations for job application failure, which were taken from a prior pilot study (Van Overwalle & Heylighen, 1995, Experiment 4) and slightly adapted to fit with the present story (see Table 2). The eight causes were selected in this pilot study because they received the highest or lowest ratings on the dimensions of locus, stability, globality and control, resulting in one single cause tapping one dimensional category (i.e., extreme). To increase the reliability of the response time measures in the present experiments, I added to each dimensional category one cause with the second highest or lowest rating from the pilot study, so that there were now two causes tapping each dimensional category (see Table 2).

To ascertain whether our subjects had roughly the same beliefs on this categorization, before the experiment, they were given a short questionnaire.

Each attribution category was described using similar wordings as in the pilot study and the present experiment. Subjects then rated the properties of the selected causes along a 6-point rating scale with the following anchors: *completely external to the candidate - completely internal to the candidate; completely temporary - completely permanent; completely specific - completely global; completely uncontrollable by the candidate - completely controllable by the candidate*. The responses were dichotomized, and the results showed that 55 out of the 56 subjects categorized at least 80 % of the causes as expected. Deletion of experimental data for which subject's dimensional judgments did not conform, had no substantial effect on the results. Hence, all data were retained and analyzed.

To control for reading time, the causes were displayed by two cue words (in capitals) as shown in Table 2. These cue words were defined by a full description of the cause which was given on a separate page just before the beginning of the experiment. The experiment started only after the subject was able to recall all the cue words and definitions verbatim.

Results

I first analyze the accuracy of the responses, and then turn to the response times. Responses that took very long were reduced to 32.8 seconds (the default maximum limit imposed by the MEL programming software) to avoid an undue influence of outliers (1.37 % of the data).²

Accuracy. The mean accuracy in function of rule and compatibility conditions is shown in the top panel of Figure 1. As can be seen, the accuracy is generally higher in the compatible condition than in the incompatible condition, as predicted. The means were analyzed in a 4 (Rule Type) x 2 (Compatibility) within-subjects MANOVA which showed the predicted main effect of compatibility, $F(1,55) = 32.59, p < .001$. Simple effect tests showed that the higher accuracy in the compatible condition was not significant for the

² The response time analyses were conducted on both correct and incorrect responses. The reason is that, due to the generally moderate accuracy, the most important ANOVAs on correct responses only would result in a substantial loss of subjects (from 56 to 52 in Experiment 1 and from 44 to 35 in Experiment 2). Moreover, the inclusion of incorrect responses represents a more conservative test of my hypotheses, because incorrect responses in an incompatible condition typically indicate that subjects did *not* take extra time to abort their natural (incorrect) responses and replace it with the prescribed response. However, an analysis of response times for correct responses revealed no substantial differences, except that in Experiment 1, the distinctiveness rule did not reach significance, $F(1,51) < 1$, which was due to large increases in the standard deviations, while the consistency rule now did reach full significance, $F(1, 51) = 5.31, p < .025$.

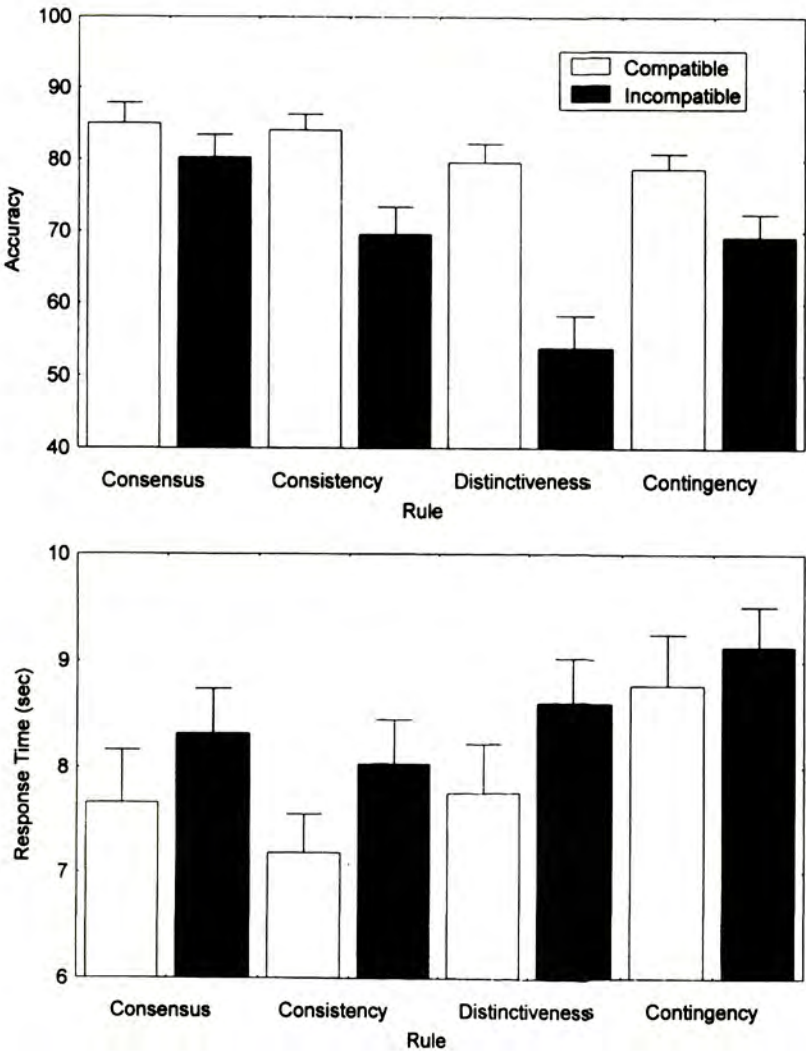


Figure 1. Accuracy in percentages (top panel) and Response Times in seconds (bottom panel) in function of Rule and Compatibility conditions: Experiment 1. Error bars depict the standard error.

consensus rule, $F(1,55) = 2.05, p = .158$; and reached statistical significance for the other three rules including the consistency rule, $F(1,55) = 12.01, p < .001$, the distinctiveness rule, $F(1,55) = 29.13, p < .001$, and the contingency rule, $F(1,55) = 9.27, p < .005$.

The MANOVA further revealed a significant main effect of rule type, $F(3,53) = 10.87, p < .001$, and a significant interaction between rule type and compatibility, $F(3,53) = 4.93, p < .004$. As can be seen in Figure 1, in the compatible condition, accuracy is uniformly high for all rules, whereas in the incompatible condition, there is much more variation between rules.

Next, I tested whether compatibility differences were found for all types of information variables and causal attributions, that is, whether they were found for both contrast and invariant information patterns (which imply the method of difference and agreement respectively), and for both contrast and invariant attributions. To do so, I conducted a 2 (Compatibility) $\times$ 2 (Information Type) $\times$ 2 (Attribution Type) within-subjects MANOVA on the mean accuracy across all four rules. The results showed that compatibility alone was highly significant, $F(1,55) = 32.59, p < .001$, and that none of the other main or interaction effects reached significance. This indicates that the accuracy under the compatible instructions was superior to the incompatible instructions regardless of type of information pattern or type of attributions, consistent with the predictions of the joint model.

Response times. The mean response times are shown in the bottom panel of Figure 1. As can be seen, subjects were generally faster in the compatible condition than in the incompatible condition, as expected. The means were analyzed in a 4 (Rule Type) $\times$ 2 (Compatibility) within-subjects MANOVA which confirmed the predicted main effect of compatibility, $F(1,55) = 11.17, p < .001$. Simple effect tests, however, showed that the difference was not significant for the consensus rule, $F(1,55) = 1.88, p = .176$, and the contingency rule, $F < 1$, marginally significant for the consistency rule, $F(1,55) = 3.93, p = .052$, and statistically significant for the distinctiveness rule, $F(1,55) = 4.40, p < .05$.

The MANOVA further revealed a significant main effect for rule type, $F(3,53) = 8.11, p < .001$. Subjects were faster for the consistency rule ($M = 7.61$ s), followed by the consensus rule ($M = 7.99$ s), and the distinctiveness rule ($M = 8.18$ s); whereas the slowest response time was for the contingency rule ($M = 8.95$ s). The interaction between rule type and compatibility condition did not reach significance, $F < 1$.

Next, I tested whether the shorter response times for the compatible rule conditions were found given both contrast and invariant types of information and attributions. A 2 (Compatibility) $\times$ 2 (Information Type) $\times$ 2 (Attribution Type) within-subjects MANOVA on the mean latencies across all four rules

revealed the expected main effect of compatibility, $F(1,55) = 11.17, p < .001$. Unlike the accuracy data, however, there was also an unexpected effect of information type, $F(1,55) = 4.48, p = .039$, which was due to the fact that compatible conditions were significantly faster than the incompatible conditions given contrastive information, $t(55) = 2.22 - 2.70, ps < .05$, but not given invariant information, $t(55) = 1.04 - 1.66, ps > .10$. Hence, the compatible instructions appeared to be applied quicker than the incompatible instructions only when contrasts were present in the covariation information, an effect which is not predicted by the joint model, but which is more consistent with the contrast model.

Experiment 2

The results of the previous experiment showed significant compatibility effects for most, but not all rules. A first potential explanation is that some of the covariation-attribution relationships are so much overlearned and routinely used that the application of its incompatible reversal does not pose excessive processing demands, such that performance remains relatively intact. This reasoning seems to provide a plausible explanation for the lack of some significant differences in the first two rules (e.g., consensus and consistency rule), because these rules attained the highest accuracy and the lowest response times overall (see Figure 1). Another potential explanation is that the internal and control response categories overlap substantially, because if the person can control his or her behavior, it is also likely that the cause is within the person. Hence, when consensus or contingency information was given, these two causal responses may have been activated together resulting in some ambiguity which may explain why the consensus and contingency rules failed to reach significant response time differences.

The second experiment was designed to explore the first explanation further. It was reasoned that if the natural information-dimension linkages are applied without much effort, then rendering the attribution task more difficult should increase the expected differences between compatible and incompatible conditions for all covariation-attribution relationships, so that they should all reach significance. This was accomplished by imposing a concurrent memory task. While working on the same attribution task as in the previous experiment, subjects additionally had to memorize an 8-digit number and had to keep it in memory as long as the rule applied. I expect that this distraction task will only slightly deteriorate response accuracy and time in the compatible conditions, because this condition is assumed to be relatively easy. In contrast, for the more difficult incompatible conditions, I expect a substantially greater deterioration in performance accuracy and time. In particular, the distraction task will leave

so little mental resources for successful execution of the incompatible instructions that accuracy will drop at chance level. Consequently, I expect that the predicted differences between compatibility conditions will increase so that conventional levels of significance will be reached for all rules.

Method

Subjects were 50 freshmen students who were drawn from the same population as the previous experiment and who participated under the same conditions. The stimulus material and procedure were identical to Experiment 1 except for the following modification: After the subjects had read the rule for a block, they got 30 seconds to memorize a random 8-digit number. They had to keep this number in memory during the whole block, until the end when they were tested for recall. In order to make the distraction task sufficiently difficult, subjects were warned not to forget the number in any case, and were advised that they should rather make some errors on the attribution task. Moreover, no card repeating the rule instructions was placed in front of the subjects. Six subjects who made mistakes on more than half of the 8-digit recall tests, were eliminated from all further analyses, leaving 44 subjects (with a 79 % correct recall on average). The responses on the questionnaire filled out before the experiment showed that all subjects categorized at least 80 % of the causes as predicted.

Results

I first analyze the accuracy of the responses, and then turn to the response times. As in the previous experiment, outliers were reduced to 32.8 seconds (0.71 % of the data).

Accuracy. The mean accuracy is shown in the top panel of Figure 2. As can be seen, the accuracy in the compatible conditions was generally higher than in the incompatible conditions, as predicted. The means were analyzed in a 4 (Rule Type) $\times$ 2 (Compatibility) within-subjects MANOVA which revealed the predicted effect of compatibility, $F(1,43) = 37.97, p < .001$. As I had hoped for, simple effect tests now revealed a significantly higher accuracy for all compatible rules, including the consensus rule $F(1,43) = 34.85, p < .001$, the consistency rule $F(1,43) = 11.02, p = .002$, the distinctiveness rule $F(1,43) = 28.78, p < .001$, and the contingency rule $F(1,43) = 7.89, p = .007$.

The MANOVA also revealed a significant main effect of rule type, $F(3,41) = 5.08, p < .004$, and a significant interaction between rule type and compatibility, $F(3,41) = 6.69, p < .001$. As can be seen in Figure 2, these latter two effects are mostly due to large variations in accuracy between rule types for the compatible

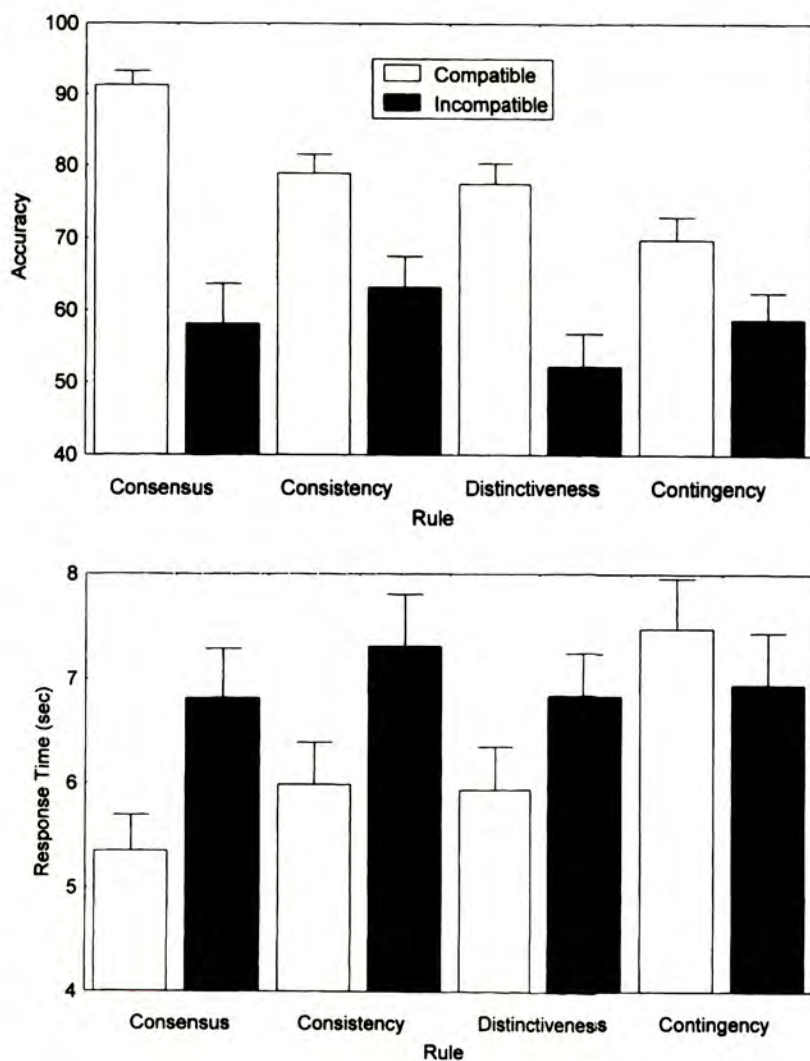


Figure 2. Accuracy in percentages (top panel) and Response Times in seconds (bottom panel) in function of Rule and Compatibility conditions: Experiment 2. Error bars depict the standard error.

condition as opposed to the incompatible condition which remained at a rather low level.

Next, I tested whether the higher accuracy for the compatible rules was found for both contrast and invariant types of information and attributions. A 2 (Compatibility) $\times$ 2 (Information Type) $\times$ 2 (Attribution Type) within-subjects MANOVA on the mean accuracy across all four rules showed a significant effect for the compatibility factor only, $F(1,43) = 37.97$, $p < .001$, indicating that the compatible condition was superior regardless of whether the covariation information or the attributions were contrastive or invariant. This is consistent with the joint model.

To test the prediction that the distraction task impaired the difficult incompatible conditions so much that accuracy dropped below chance, I conducted Binomial tests for each rule in each compatibility condition. As predicted, the accuracy was not above 50 % chance level for all incompatible rules ($ps > .05$, one-tailed), whereas it remained above chance level for all compatible rules ($ps < .01$, one-tailed).

Response times. The mean response times are shown in the bottom panel of Figure 2. On average, the responses were over one second quicker than in the previous experiment, which may indicate that subjects in this experiment simply worked faster in order not to forget the number. As can be seen, the predicted compatibility effect was found for the first three rules (as all show faster response times in the compatible condition). However, contrary to the predictions, there was a reversal for the contingency rule which, as we will see shortly, was not significant.

The means were analyzed in a 4 (Rule Type) $\times$ 2 (Compatibility) within-subjects MANOVA which revealed the predicted main effect of compatibility, $F(1,43) = 10.72$, $p < .002$. Simple effect tests revealed that subjects were significantly faster in the compatible condition of the first three rules, including the consensus rule $F(1,43) = 8.46$, $p = .006$, the consistency rule $F(1,43) = 9.35$, $p = .004$, and the distinctiveness rule $F(1,43) = 4.48$, $p = .040$; whereas the contingency rule failed to reach significance, $F < 1$.

The MANOVA further revealed a marginally significant main effect for rule type, $F(3,41) = 2.60$, $p = .065$. Overall, subjects tended to be faster given the consensus rule ($M = 6.08$ s), followed by the distinctiveness rule ($M = 6.38$ s), and the consistency rule ($M = 6.65$ s); whereas the contingency rule tended to be slowest ($M = 7.21$ s). The interaction between rule type and compatibility was also marginally significant, $F(3,41) = 2.73$, $p = .056$, which was most probably due to the unexpected reversal observed for the contingency rule.

Next, I tested whether the faster response times for the compatible version of the first three rules were revealed for both contrast and invariant types of information and attributions. A 2 (Compatibility) $\times$ 2 (Information Type) $\times$ 2

(Attribution Type) within-subjects MANOVA on the mean latencies of the first three rules (without the contingency rule) revealed the expected main effect of compatibility only, $F(1,43) = 18.17, p < .001$. This indicates that unlike Experiment 1, the three compatible rules lead to quicker responses regardless of the type of information variables or attributions. This is again in line with the joint model.

Taken together, the results of the second experiment showed more substantial and more consistent differences between the rule conditions than in the first experiment, which supports the prediction that the distraction task impaired correct and fast execution of the relatively difficult incompatible conditions even more. Moreover, as predicted by the joint model, these differences were equally strong regardless of whether the covariation information implied a contrast or invariant pattern, or whether the attribution involved a contrast or invariant cause. Nevertheless, there was one unexpected finding in that the response times of the contingency rule were not significantly different between compatibility conditions, although the accuracy data for this rule were, as expected, significantly better in the compatibility condition.

General Discussion

Let me first summarize the main results of the two experiments. It was predicted that imposing a causal judgment rule incompatible with subjects' habitual responses, would increase the number of errors and the response time in comparison with a rule that is compatible with subjects' natural causal inferences. Conform to these predictions, in the first experiment, the results showed that the responses in the incompatible conditions were less accurate and slower than the compatible conditions for all rules, although some differences were only marginally significant and the response time difference for the contingency rule was far below significance. To obtain sharper differences, in Experiment 2, I imposed an additional distraction task during the attributional judgment task. The results now showed significantly lower accuracy and slower responses for all incompatible rules, although the response times for the contingency rule again failed to reach significance. In sum, the predictions were strongly confirmed for three covariation-attribution relationships proposed by Van Overwalle and Heylighen's (1995) joint model, whereas the contingency-control relationship was supported only by the accuracy data. It is instructive to note that in their 1995 studies, Van Overwalle and Heylighen also failed to obtain strong evidence for this latter relationship. Taken together, the whole set of results suggests that the relationship between contingency and control is not very unequivocal.

As noted in the introduction, a possible explanation for this failure is that

although low contingency strongly suggests that the actor had no control over his or her outcomes, the reverse pattern of high contingency does not necessarily imply that the actor did have full control. To infer control, perceivers typically assume that the actor also had foreknowledge and intention. As Heider (1958) noted, a person "is considered responsible, directly or indirectly, for any aftereffect he may have foreseen" (p. 113). If such foreknowledge or intention is not present, perceivers may assume that the outcome changed simply through some trial and error, lucky guesses or other random behaviors which do not involve control by the actor. Another, related explanation is that if persons have control over their behavior, then it is very likely that the cause also resides within the person. Hence, there may be some ambiguity in the attributional category (internal or controllable?) that is most strongly activated when contingency information is provided. In sum, it appears that the action-contingency information is perhaps less fundamental than the other covariation variables proposed by Kelley (1967), and further research is certainly needed to establish the importance of this information in determining causal controllability.

Although the results were largely consistent with my predictions, the implications drawn from this research need some qualification because the present stimulus material (1) was limited to one social event, (2) was depicted in a verbal summary information, and (3) was presented together with a particular rule how to infer the cause. However, not all of these limitations are equally problematic for the basic predictions of the joint model.

First, the fact that the stimulus material was limited to one event (i.e., rejection of a job applicant) was necessitated by the design of my experiments in which I used concrete causal explanations which tapped eight causal categories. It would be an enormous task to select causal explanations with sufficiently strong and consensual causal properties for a large series of events. Irrespective of this limitation, other data by Van Overwalle and Heylighen (1995) and Van Overwalle (1997a) supported the joint model's predictions for a much larger set of outcomes, including other social and achievement events. Moreover, there is strong evidence that Mill's methods of difference and agreement apply to other kinds of attributional judgments, such as judgments about a person's dispositions. For instance, Hilton et al. (1995) and Van Overwalle (1997b) demonstrated that a person disposition implies that the target person has a particular property more than the average person (i.e., low consensus requiring the method of difference) and displays this property across many situations and occasions (i.e., low distinctiveness and high consistency requiring the method of agreement).

A second limitation of the present experiments is that covariation information was presented in a verbal summary format. Although summary information typically reflects how attributions are made on the basis of social communication,

it does reveal very little about incidental day-by-day causal inferences. It is unclear to what extent such case-by-case data would make the relevant dimensions more or less salient, or the underlying methods of causality more or less potent. This is a shortcoming not only of the present work, but of most attributional research where the same paradigm has been commonly used (e.g., Cheng & Novick, 1990; Försterling, 1989; Hilton & Slugoski, 1986; Hilton, Smith, & Kim, 1994; Hewstone & Jaspars, 1987).

A third concern is that imposing a particular causal rule begs the question whether the participants would ever use such a rule spontaneously. There is a long tradition of research showing that people employ the method of difference in making causal attributions (e.g., Cheng & Novick, 1990; Försterling, 1989; Hilton & Slugoski, 1986), but research on the method of agreement is only beginning to emerge (Hilton, Smith, & Kim, 1995; Van Overwalle & Heylighen, 1995; Van Overwalle, 1997a, 1997b). When the task provides only invariant information (e.g., all candidates failed), do people then settle for an indistinct invariant categorical explanation (e.g., external cause) using the method of agreement, or do they search for more detailed causes by applying the method of difference on events outside the focal information? Although the joint model does not deny this latter possibility, it predicts that invariant categorical explanations *can* be sufficient on their own. Thus, although the method of agreement does not allow perceivers to select with certainty a particular cause, it enables them to limit the causal field to invariant categories, where they may or may not search further for plausible causal candidates (see also Mill, 1872/1973).

In contrast, Cheng and Novick (1990, 1991) would take a much stronger position and argue that given invariant information, people do not make any attributions at all, *unless* they can invoke default assumptions from prior knowledge which contrasts against the focal information. As noted earlier, this strong assumption leads to the prediction that the compatibility effect should typically be weaker or absent given invariant information patterns and attributions. However, such differences were generally not found in my data.

However, the design of the present experiments may perhaps not allow for a clear crucial test between the two theories. The causes used were selected from an earlier pilot study (Van Overwalle & Heylighen, 1995) in which subjects had to come up with plausible explanations. This procedure seems to assure that the explanations were viewed as potentially causal by the majority of the subjects, and not as irrelevant. If this was the case, then Cheng and Novick's (1990) model would predict a smaller difference between causal and enabling conditions (depending on whether the method of difference or agreement was applied respectively). However, this rather weak difference in causal status was not explicitly measured in our experiments and could therefore have been left undetected.

Nevertheless, earlier research in which subjects had to judge broad attributional categories has documented that they select both contrast and invariant categories in line with the methods of difference and agreement, although subjects tend to use more contrast categories overall (Van Overwalle & Heylighen, 1995; Van Overwalle, 1997a). Taken together, the results of these and the present studies suggest that invariant categories and causes are used quite often. Hence, both the method of difference and the method of agreement, as exemplified in the joint model, seem to guide causal judgments.

References

- Abramson, L. V., Seligman, M. E. P., & Teasdale, J. D. (1978). Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology*, 87, 49-74.
- Bradbury, T. N., & Fincham, F. D. (1990). Attributions in marriage: Review and critique. *Psychological Bulletin*, 107, 3-33.
- Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology*, 58, 545-567.
- Cheng, P. W., & Novick, L. R. (1991). Causes versus enabling conditions. *Cognition*, 40, 83-120.
- de Jong, P. F., Koomen, W., & Mellenbergh, G. J. (1988). Structure of causes for success and failure: A multidimensional scaling analysis of preference judgments. *Journal of Personality and Social Psychology*, 55, 718-725.
- Folkes, V. S. (1984). Consumer reactions to product failure: An attributional approach. *Journal of Consumer Research*, 11, 398-409.
- Försterling, F. (1989) Models of covariation and attribution: How do they relate to the analogy of analysis of variance? *Journal of Personality and Social Psychology*, 57, 615-625.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Hewstone, M., & Jaspars, J. (1987). Covariation and causal attribution: A logical model of the intuitive analysis of variance. *Journal of Personality and Social Psychology*, 53, 663-673.
- Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review*, 93, 75-88.
- Hilton, D. J., Smith, R. H., & Kim S. H. (1995). The process of causal explanation and dispositional attribution. *Journal of Personality and Social Psychology*, 68, 377-387.
- Kelley, H. H. (1967). Attribution in social psychology. *Nebraska Symposium on Motivation*, 15, 192-238.
- Kornblum, S., Hasbrouc, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility—A model and taxonomy. *Psychological Review*, 97, 253-270.
- Mill, J. S. (1872/1973). A system of logic ratiocinative and inductive. In J. M. Robson (Ed.), *Collected works of John Stuart Mill* (Vols. 7-8). Toronto, Canada: University

- of Toronto Press.
- Peterson, C., & Seligman, M. E. P. (1984). Causal explanations as a risk factor for depression: Theory and evidence. *Psychological Review*, 91, 347-374.
- Van Overwalle, F. (1989). Structure of freshmen's causal attributions for exam performance. *Journal of Educational Psychology*, 81, 400-407.
- Van Overwalle, F. (1997a) A test of the Joint model of causal attribution. *European Journal of Social Psychology*, 221-236.
- Van Overwalle, F. (1997b) Dispositional attributions require the joint methods of difference and agreement. *Personality and Social Psychology Bulletin*, 23, 974-980.
- Van Overwalle, F., & Heylighen, F. (1995). Relating covariation information to causal dimensions through principles of contrast and invariance. *European Journal of Social Psychology*, 25, 435-455.
- Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer-Verlag.



Received June, 1997
Revision received October, 1997
Accepted, January 1998