

Generating explanations of social and nonsocial events

KATHLEEN M. GALOTTI
Carleton College, Northfield, Minnesota

and

DEBRA A. KOSSMAN and JOHN P. SABINI
University of Pennsylvania, Philadelphia, Pennsylvania

Thirty-six undergraduates generated explanations for five social events (e.g., "A patron leaves a small tip in a restaurant") and five nonsocial events (e.g., "A block of a main city street has been closed to traffic"). The number of distinct explanations, the number of different types of explanations, the originality and elaboration of the explanations, and the systematicity of generating explanations were scored. Initial results suggested more thoughtful and less stereotypic performance when the subjects generated explanations for social events than for nonsocial events. Later analyses showed that many of these results depended on the particular stimulus items. Measures for describing and evaluating reasoning in everyday contexts are discussed.

How is the structure of everyday reasoning to be characterized? To answer this question, one must first identify appropriate everyday reasoning tasks. An everyday reasoning task should make demands of people that mimic those encountered in the normal course of life. One such task is generating possible explanations for events.

Casual reflection suggests that people often spontaneously spend time and effort considering alternative causes for various kinds of events. Indeed, attribution theorists from Heider (1958) on have (often exclusively) construed social cognition as reasoning about causal explanation. Findings from some studies in the attribution literature parallel those from the everyday reasoning literature. Hansen (1980), for example, suggested that people hold "naive causal hypotheses" and use them as a guide for seeking information to test their attributions. In so doing, however, they exhibit a confirmation bias, searching only until a sufficient cause is found. This account is reminiscent of that presented by Perkins, Allen, and Hafner (1983), who argued that people typically fail to move beyond an account that makes superficial sense when reasoning about a political issue.

In recent studies of everyday reasoning, Perkins and associates (Perkins, 1985; Perkins et al., 1983) argued that the average person generally does a poor job when reasoning about the truth of a claim. The major limitation of thinking, Perkins asserted, is a failure to examine initial hypotheses critically. Indeed, many existing models

of reasoning performance (both laboratory and everyday) rely on the following principle: Good reasoning comes about in large part as the result of considering alternative possibilities (Baron, 1985; Galotti, Baron, & Sabini, 1986; Johnson-Laird, 1983). People who consider only one possibility are less likely to come to the correct conclusion, if one exists, or to a high-quality solution, when many possible solutions exist. Those who consider multiple possibilities have a greater chance of overcoming biases, noticing contradictions, and examining unwarranted assumptions.

The goal of the present study was to characterize the reasoning used in generating explanations. A central question was whether everyday reasoning would be subject to the dramatic content effects found with formal reasoning tasks, such as syllogistic or propositional reasoning (Revlín & Mayer, 1978). A second question was whether reasoning performance would differ as a function of whether the event being reasoned about is primarily social or nonsocial in nature.

Shultz (1981) argued that social causal reasoning is similar to nonsocial causal reasoning but also differs from it in significant ways. He noted that reasoning about any kind of event relies upon the idea that causes actually generate or produce their effects. However, social causal reasoning concerns itself with the actor's enactment of intentions as the type of generative mechanism, whereas physical causal reasoning concerns itself with the transmission of energy as the generative mechanism. Presumably, different generative mechanisms give rise to different kinds of causal reasoning. However, thus far the claims made about differences in kinds of causal reasoning have all been in terms of content.

Our question, in contrast, was whether the *process* and/or the *structure* of reasoning would be affected by

We thank Andrea Matchett and Paula Erickson for their help in analyzing the data. We are also grateful to Jon Baron, Lloyd Komatsu, Steven Kozberg, and Andrea Matchett for their comments on an earlier draft. Correspondence may be addressed to Kathleen M. Galotti, Department of Psychology, Carleton College, One North College Street, Northfield, MN 55057-4025.

the type of event (social or nonsocial) under consideration. One might argue, for example, that causal reasoning about social events is inherently more interesting, engaging, and/or familiar to most people and thus elicits more thorough or creative performance. Alternatively, perhaps nonsocial events provoke less emotional involvement and, hence, less biased reasoning, manifested (for example) by a more systematic consideration of possibilities.

METHOD

Subjects

Thirty-six University of Pennsylvania students were recruited from psychology classes to serve as subjects. They were paid \$6 for participating in a 90-min session.

Materials

The subjects received two booklets. The social events booklet contained five one-sentence descriptions (one per page) of an interaction between an actor and one or more others. The descriptions were: "A patron leaves a small tip for a waitress," "A supervisor suggests that an employee be promoted," "Two friends, A and B, had plans to go to the movies, but A cancels," "A professor grades a paper with an 'A,'" and "A voter votes to increase property taxes." For each description, the subjects were instructed to list as many distinct explanations as they could in the time allotted (15 min for all five items). The nonsocial events booklet was similar in format to the social events booklet, but contained five one-sentence descriptions (one per page) of an event in which the focus was not the interaction between individuals. The descriptions were: "The eleven o'clock news came on later than eleven o'clock," "A person is diagnosed as having a certain disease, but doesn't improve when prescribed the medication that works for most people," "Milk in a refrigerator has spoiled," "A block of a main city street has been closed to traffic," and "A burglar is able to enter a house and burglarize it while a family sleeps within, and while the front door is locked."

Procedure

The subjects were run in groups of up to 12, and were allotted 15 min for each booklet. The order of booklets was counterbalanced.

RESULTS

Dependent Measures

Responses were scored for several content-independent measures. These measures were adapted from the divergent thinking literature (Torrance, 1972, 1974), and are used to compare people's productions of explanations across different types of items. *Fluency* is the number of distinct explanations generated. *Flexibility* is the number of different types of responses generated. *Originality* describes the statistical unusualness of the explanations a subject generates. *Elaboration* is the average amount of "extra detail" added to individual explanations. *Category switches* reflects the manner in which a subject generates a set of responses; it is a tally of the number of times that a subject switches from one flexibility category to another.

Presumably, careful thinking about the cause(s) of an event is reflected by considering a large number of explanations (high fluency), explanations of different sorts (high flexibility), and by moving beyond the most obvious explanations (high originality). It is possible that elaborating one or two explanations results in the abro-

gation of the list of other explanations that might have been thought of (see Perkins, 1985), hence the inclusion of the elaboration measure. Finally, the number of category switches is intended to reflect the "orderliness" with which a subject considers various explanations. A very methodical subject might be expected to generate all the explanations he or she can think of in one category before moving on to the next, and would thus switch categories relatively infrequently.

Calculation of these five structural measures followed along lines similar to those of Torrance (1972). First, for each item, a list of all of the distinct responses from the entire subject pool was compiled, along with a frequency tally (from all subjects) for each response. A subject's fluency score was the number of responses on the list that the subject gave for that item. Computation of a subject's originality score for an item was more complicated. First, the frequency tally for each response to the item was converted to a proportion. For example, if 18 of the 36 subjects gave a particular response, the score for that response was .50. For any given subject's set of responses to an item, the mean of the logarithms of the responses' scores was calculated. The absolute value of this mean constituted the originality score for that item. A logarithmic transformation was used to give greater weight to relatively infrequent responses, and less weight to relatively frequent ones, following our intuitions that very infrequent responses ought to contribute more heavily to an originality measure. Because logarithms of proportions result in negative numbers, using absolute values makes the measure again correspond to intuition: higher originality scores reflect the generation of less frequent responses.

To calculate flexibility and category-switching scores, responses for each item were categorized (by the authors and two research assistants unfamiliar with the purposes of the study). After the final taxonomies were developed, the first author and a research assistant independently coded all responses on each list according to the final taxonomy for each item. A subject's flexibility score for a particular item was the number of different categories in that item's taxonomy that his or her explanations fell into. Median interrater reliabilities for the categories for the 10 items ranged from .95 to 1.00; median interrater reliabilities using the more conservative κ ranged from .66 to 1.00. A subject's category-switching score was simply the number of times, within a set of responses to a single item, a subject switched from one flexibility category to another, minus the flexibility score for that item (because if a subject uses n flexibility categories in responding to a given item, he or she must switch categories at least $n - 1$ times). Note that this measure is bounded from above by the fluency score for a given item.

To calculate elaboration measures, the three authors independently rated all of the responses according to how much each response was elaborated. Responses that conveyed what was felt to be only "basic" information received a score of 1; higher scores were given for each elaboration over and above "basic" information. For each

Table 1
Intercorrelations of Structural Measures by Type

Item Type	Structural Measures									
	Social					Nonsocial				
	Fluency (1)	Flexibility (2)	Originality (3)	Elaboration (4)	Mean Category Switches (5)	Fluency (6)	Flexibility (7)	Originality (8)	Elaboration (9)	Mean Category Switches (10)
(2)		.97*								
(3)		.47†	.47†							
(4)		-.05	-.01	.27§						
(5)		.79*	.71*	.29‡	-.20§					
(6)		.81*	.82*	.38‡	.00	.65*				
(7)		.80*	.80*	.45†	.03	.66*	.86*			
(8)		.67*	.68*	.55*	.15	.56*	.75*	.74*		
(9)		.12	.17	.34‡	.60*	-.16	.04	.23§	.33‡	
(10)		.72*	.73*	.37‡	-.16	.59*	.80*	.68*	.64*	.00

Note—Boldface items indicate corresponding measures in the social and nonsocial domains (e.g., social fluency and non-social fluency). * $p < .001$. † $p < .01$. ‡ $p < .05$. § $p < .10$.

additional proposition incorporated into a response, one point was added to the elaboration score for that response. The elaboration for each response was the mean of the ratings given by the three authors (the median of the interrater reliabilities, computed by Pearson product-moment correlations, was .62). A subject's elaboration score for a particular item was the mean of the elaboration ratings for all of the responses given to that item.

Internal reliabilities across all 10 items (using coefficient α) were .92 for fluency, .87 for flexibility, .75 for originality, .66 for elaboration, and .74 for category switches.

Performance as a Function of Item Type

Means of the five structural measures across the five social items and across the five nonsocial items were calculated. Table 1 presents the intercorrelations for the five measures, within and across domains (social vs. non-

social). Table 1 also presents the intercorrelations of the five measures by item type. It shows that all of the measures were highly correlated across the two types of items. Fluency, flexibility, and category switches were also highly intercorrelated, and elaboration was generally uncorrelated with other measures.

Differences in these measures as a function of the type of item—social or nonsocial—were examined by ANOVAs on the collapsed means (social vs. nonsocial) for each of the five structural measures (ANOVAs were used rather than t tests to facilitate calculation of the min F' statistic, discussed below). The results, shown in Table 2, initially suggested that the subjects exhibited significantly higher flexibility, originality, and elaboration, and made significantly more category switches with social items. However, to test whether these results would also generalize simultaneously to another sample of items, we calculated min F' , a lower bound of the appropriate

Table 2
Structural Measures by Item Within Type

Measure	Item and Type											
	Nonsocial						Social					
	News	Refr	Traf	Burg	Dis	M	Tip	Movi	Grad	Prom	Tax	M
Fluency	5.80	5.74	8.31	6.71	5.74	6.52	7.09	8.11	6.06	5.92	5.17	6.48
	$F(1,35) = 0.04$, n.s.; min $F'(1,10) = 0.004$, n.s.											
Flexibility	3.69	3.86	5.11	4.08	5.22	4.39	6.00	6.39	5.17	5.03	4.33	5.38
	$F(1,35) = 52.58$, $p < .001$; min $F'(1,9) = 3.84$, n.s.											
Originality	1.69	1.64	1.88	1.95	1.45	1.72	1.61	1.92	2.30	2.14	2.58	2.11
	$F(1,35) = 71.96$, $p < .001$; min $F'(1,9) = 4.29$, $p < .10$											
Elaboration	1.12	1.17	1.13	1.16	1.07	1.13	1.26	1.15	1.20	1.26	1.26	1.23
	$F(1,35) = 37.11$, $p < .001$; min $F'(1,13) = 8.74$, $p < .05$											
Category Switches	-0.33	-0.14	0.42	0.00	-0.61	-0.47	-0.39	0.00	-0.58	-0.67	-0.75	-0.13
	$F(1,35) = 16.89$, $p < .001$; min $F'(1,11) = 2.20$, n.s.											

Note—News = news, Refr = refrigerator, Traf = traffic, Burg = burglar, Dis = disease, Tip = tip, Movi = movie, Grad = grade, Prom = promo, Tax = taxes. See text for a complete listing of the items.

quasi- F ratio, which treats both subjects and items as random effects. (For more on the calculation of min F' , and where it is appropriate, see Clark, 1973.) Table 2 shows that for all of the measures, except elaboration, the results seem to be specific to the particular items used. Table 2 also presents the means, by item, of all of the structural measures, and reveals striking variation by item.

DISCUSSION

The quality of the subjects' reasoning about the causes of social events seemed at first glance superior to their reasoning about the causes of nonsocial events. Although no overall differences were found in the number of distinct explanations generated (fluency), there were significant differences that suggested that the subjects generated a wider variety of explanations (higher flexibility), more original explanations (higher originality), more detailed explanations (higher elaboration), and generated explanations less systematically (fewer category switches) for social events. However, when the stimulus items and subjects were treated as simultaneous random effects in a subsequent analysis, the social/nonsocial distinction was reliable only for the elaboration measure.

This lack of reliable differences suggests, in turn, the presence of content effects in everyday reasoning, reminiscent of the content effects commonly found in traditional laboratory studies of reasoning. However, inspection of Table 2 does not reveal any interpretable pattern to the content effect. One might suppose that familiar items would elicit the best reasoning, because more familiar items should be easier to encode. One might further suppose that the "grade," "movie," and "tip" items would be the most familiar to our undergraduate subjects, and that the "tax" item would be the least familiar. However, the first three items did not elicit better performance than the presumably less familiar item.

The pattern of intercorrelation of the structural measures (see Table 1), which is evident for both social and nonsocial items, suggests a number of things about performance in generating explanations. First, fluency and flexibility are highly correlated, which suggests that sheer persistence can bring about the generation of different kinds of explanations. Obviously, this account is speculative, since correlation does not show directionality. Second, originality is highly correlated with both fluency and flexibility, which suggests that persistence in generation can lead to less stereotypic thinking. Again, the above caveat applies. Third, category switching rises with fluency as well, which suggests that as one continues to generate explanations, one may become less systematic. Finally, and contrary to expectation, elaboration was not negatively

related to fluency. This suggests that cognitively investing in one explanation (i.e., adding details) does not preclude or reduce the generation of other alternatives.

More direct tests of the variables that influence how people generate explanations are needed. For example, does persistence alone necessarily lead to greater flexibility and/or originality? Future investigation should also probe the similarities and differences of explanation generation to other forms of everyday reasoning, such as personal decision making (Nisbett & Ross, 1980) or argument evaluation (Perkins, 1985).

REFERENCES

- BARON, J. (1985). *Rationality and intelligence*. Cambridge: Cambridge University Press.
- CLARK, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning & Verbal Behavior*, *12*, 335-359.
- GALOTTI, K. M., BARON, J., & SABINI, J. P. (1986). Individual differences in syllogistic reasoning: Deduction rules or mental models? *Journal of Experimental Psychology: General*, *115*, 16-25.
- HANSEN, R. D. (1980). Commonsense attribution. *Journal of Personality & Social Psychology*, *39*, 996-1009.
- HEIDER, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- JOHNSON-LAIRD, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.
- NISBETT, R., & ROSS, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- PERKINS, D. N. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology*, *77*, 562-571.
- PERKINS, D. N., ALLEN, R., & HAFNER, J. (1983). Difficulties in everyday reasoning. In W. Maxwell (Ed.), *Thinking: The expanding frontier* (pp. 177-189). Philadelphia: Franklin Institute.
- REVLIN, R., & MAYER, R. E. (1978). *Human reasoning*. New York: Holt.
- SHULTZ, T. R. (1981). Causal reasoning in the social and nonsocial realms. *Canadian Journal of Behavioral Science*, *14*, 307-322.
- TORRANCE, E. P. (1972). Can we teach children to think creatively? *Journal of Creative Behavior*, *6*, 114-143.
- TORRANCE, E. P. (1974). *Torrance tests of creative thinking*. Bensenville, IL: Scholastic Testing Service.

(Manuscript received January 5, 1990.)