Cognition, 14 (1983) 237–273

The psychological causality implicit in language

ROGER BROWN*
DEBORAH FISH
Harvard University

Abstract

We are concerned with the causality implicit in English verbs that name interactions, either mental or behavioral, between two persons, verbs such as like, notice *(mental), and* help, cheat *(behavioral) in such a context as* Ted — Paul. *Using four different methods, we show that adult native speakers think of*

*Preparation of this article was supported in part by Research Grant No. BNS-82-09378 from the National Science Foundation to Roger Brown. For reading the paper and making valuable suggestions we would like to thank Joan Bresnan, Jerome S. Bruner, William Cooper, Catherine Garvey, Fritz Heider, Harold H. Kelley, Michael Maratsos, Kathleen McCartney, Stanley Milgram, George Miller, Miriam Schustack, W.V. Quine and Eric Wanner. Requests for reprints should be sent to Roger Brown, Department of Psychology and Social Relations, Harvard University, 1270 William James Hall, Cambridge, Massachusetts 02138.

Soon after completing this paper, we discovered that we had been anticipated in the study of implicit causality by Garvey, Caramazza and their associates (1974, 1976) and by Caramazza *et al.* (1977). This prior work uses the very term 'implicit causality' and suggests, as we do, that certain verbs carry a semantic feature of causal attribution that is probably not binary in nature but is better conceived of as a continuum. The starting point of the Garvey and Caramazza studies was not, as was ours, derivational adjectives, but rather the assignment of pronoun antecedents when sentence context is not determinative. In one study (Garvey, Caramazza, and Yates, 1976), subjects were given sentences with the abstract format: *NP1 Past-V NP2 because Pro...* (e.g. *John admired Paul because he...*). The pronoun *he* is ambiguous in the example since both NP1 and NP2 are masculine nouns. However, subjects in completing the sentence (e.g., *was a fine athlete*) revealed which noun they took to be co-referential with the pronoun (in this case, *Paul*). Finding that subjects manifested high agreement in the disambiguation of some verbs, the authors suggested that a causal attribution feature in the verb root selected between NP1 and NP2. In another study (Caramazza *et al.*, 1977), it was shown that the time to process sentences of the type described above depended on whether the semantic bias of the subordinate clause was congruent with the inherent bias of the verb (as in *Rose envied Mary because she always looked so neat*) or incongruent with the inherent bias (as in *Anne envied Vicki because she never looked as neat*). In part, the results of the Garvey and Caramazza studies appear to be consistent with the theory of the two schemas and, in fact, the disambiguation of ambiguous pronouns may provide very strong evidence for the existence of the schemas. Compare *Rose admires Mary because she is so unconventional* (*she* is *Mary*) and *Rose charms Mary because she is so unconventional* (*she* is *Rose*). However, there are some results in the Garvey and Caramazza studies which the two schemas, as presently formulated, cannot accommodate, and there are many for which the theory simply makes no prediction. For example, in sentences in which the referents of NP1 and NP2 have unequal status (e.g., *The father scolded his son because he...*), the inherent causality of the verb must be overwhelmed by real-world probabilities. The Garvey and Caramazza sentences were deliberately designed sometimes to use nouns referring to persons of unequal status and were deliberately varied in other ways, which makes it impossible to decide in many cases whether or not the results have any bearing on the two schemas. In subsequent reports on our research we plan to discuss more fully the prior discoveries of Garvey and Caramazza.

0010-0277/83/$11.60

causality in such verbs as unequally apportioned between interactants. For behavioral (or action) verbs greater causal weight is given to the Agent argument of the verb (e.g., Ted *in* Ted helps Paul*) than to the Patient argument (*Paul*). For mental (or state) verbs greater causal weight is given to the Stimulus argument of the verb (e.g.,* Paul *in* Ted likes Paul*) than to the Experiencer argument (*Ted*). For English verbs of the type studied, derivational adjectives often exist (e.g.,* helpful, cheating, likable, noticeable*). Such adjectives are attributive to one or the other argument of the verb base (Agent or Patient; Stimulus or Experiencer). We show that the direction of causal attribution in the adjective (e.g.,* helpful *is attributive to* Ted *the Agent;* likable *is attributive to* Paul *the Stimulus) predicts the primary causal weightings assigned in our experimental tasks. We also show that in the English language adjectives derived from action verbs are almost attributive to the Agent and adjectives derived from state verbs to the Stimulus. Because certain facts about English morphology predict certain ways of thinking about causality, our main finding may seem to be a Whorfian one, a demonstration that language affects thought. However, we argue that it is not that but rather a demonstration that two modes of thought (the Agent–Patient Schema and the Stimulus–Experiencer Schema) affect language use. We suggest that the schemas are universals of human thought.*

It was Fritz Heider (1958) who awakened psychological interest in the attribution of causes in everyday social life. By 1979 (Kelley and Michela, 1979), more than 1,000 experimental and theoretical papers had been published on this subject. Earlier work concentrated on the rules of inductive inference employed and, especially, on Harold Kelley's (1967) formulation of these rules in terms of the ANOVA cube. Later work has concentrated on the formulation of such general biases and errors as the Fundamental Attribution Error (Ross, 1977); the Actor–Observer Divergence (Jones and Nisbett, 1972); and the Self-Serving Bias (Ross, 1977). The present paper sets forth evidence that a theory of psychological causality is implicit in natural language itself. This theory, prior to all general biases and errors, might be learned by the inductive calculus from certain facts about human experience and action and might exert a universal influence on human languages. In a way, a new way, this paper returns to Heider's own starting point—the analysis of the vernacular.

Overview of the argument

Research begins with a paper-and-pencil task that subjects find quite unusually boring and mindless, a task evoking responses which subjects believe to be entirely random. Subjects are mistaken on all counts.

In skeletal form the task[1] is this:

> *Ted likes Paul.*
> How likely is it that this is because:
> A. Ted is the kind of person that likes people.
> Not likely 1 2 3 4 5 6 7 8 9 Definitely likely
> B. Paul is the kind of person that people like.
> Not likely 1 2 3 4 5 6 7 8 9 Definitely likely
> C. Some other reason.
> Not likely 1 2 3 4 5 6 7 8 9 Definitely likely
>
> You are asked to rate each of these explanations on how likely it is to have caused the stated event. Do so by circling the appropriate numbers. Please make the rating of each cause (A, B, and C) independent of the other causes; i.e., the three circled numbers are not required to sum to any particular value. Please also make the ratings of each sentence-event independent of each other sentence-event. The same proper names (Paul and Ted) are used throughout and are not intended to have any significance. You should not, for instance, think of Paul (or Ted) as one person who should be rated in a consistent way from one sentence-event to another. The names are arbitrary and have been randomly permuted in sentence positions.

What varies from problem to problem (page to page) is the verb. Interest centers on Alternatives A and B, and the question is whether the greater causal weight in the event named by the sentence is assigned to the subject of the sentence (here *Ted*) or to the object (here *Paul*). In one study or another we have, to date, used 60 verbs, all transitive, and all acceptable in the context Ted — Paul. The list includes *help, harm, disobey, charm, attract, deceive, like, honor,* and *detest.*

With materials of the kind described, we have made a discovery that we think remarkable: the causal balance in the verb is predicted by a fact about English derivational morphology. To be sure, the result obtained is the result predicted but, before the fact, we considered it a very low probability outcome. All of the verbs used form bases for derived adjectives: *help – helpful; compete – competitive; charm – charming; attract – attractive; like – likable; detest – detestable;* etc. If we take the depicting sentence (e.g., *Ted helps*

[1]This task is taken from Cunningham, Starr, and Kanouse (1979).

Paul; *Ted likes Paul*) and ask the causal question, *Why?*, then acceptable minimal answers are provided by the respective derived adjectives. Because *Ted is helpful* and because *Paul is likable*. Between the two adjectives in this example there is an important difference; *helpful* is attributive to, or predicative of, Ted, the sentence subject, whereas *likable* is attributive to *Paul*, the sentence object. English seems not to have a dispositional adjective based on *help* attributive to the sentence object (the person helped); we have no form *helpable*. English seems also not to have an adjective based on *like* that is attributive to the subject (the person who likes; we have neither *likeful* nor *liking* as an adjective).

Limiting ourselves, for expository purposes, to just 12 verbs (with full evidence to come), we find the pattern of Table 1. Consider the first sentence: *Ted helps Paul*. When an experimental subject assigned a higher rating to Alternative A ('Ted is the kind of person who helps people') than to Alternative B ('Paul is the kind of person that people help'), this is considered to be an outcome of the type 'S > O' since *Ted* is the subject (S) of the sentence and *Paul* is the object (O). Other possible outcomes are, therefore, 'O > S' and 'S = O'. For the eight verbs of Table 1(a), large majorities of experimental subjects (n = 20) gave responses of the type 'S > O'. This is to say that

Table 1. *Characteristic results for verbs forming derivational adjectives attributive to the sentence subject* (S) *and verbs forming adjectives attributive to the sentence object* (O) (n = 20)

(a) Verbs which form derivational adjectives attributive to the subject (S)

Sentence	Derived Adjective		S > O	O > S	S = O
Ted helps Paul	*helpful*	[*helpable*?]	15	0	5
Ted cheats Paul	*cheating*	[*cheatable*?]	17	0	3
Ted competes with Paul	*competitive*	[*competitable with*?]	13	1	6
Ted criticizes Paul	*critical*	[*criticizeable*?]	15	0	5
Ted attracts Paul	*attractive*	[*attractable*?]	15	2	3
Ted charms Paul	*charming*	[*charmable*?]	18	1	1
Ted deceives Paul	*deceptive*	[*deceivable*?]	15	1	4
Ted troubles Paul	*troublesome*	[*troublable*?]	15	0	5

(b) Verbs which form derivational adjectives attributive to the object (O)

Ted likes Paul	*likable*	[*liker*; *-ing*; *-ful*?]	1	13	6
Ted detests Paul	*detestable*	[*detester*; *-ing*; *-ful*?]	1	16	3
Ted loathes Paul	*loathsome*	[*loather*; *-ing*; *-ful*?]	2	17	1
Ted notices Paul	*noticeable*	[*noticer*; *-ing*; *-ful*?]	1	15	4

interpersonal events involving help, cheating, competition, criticism, attraction, charm, deception, and trouble were primarily attributed to a general disposition in him who helps, cheats, competes, criticizes, attracts, charms, deceives, and troubles rather than to him who is helped, cheated, competed-with, criticized, attracted, charmed, deceived, and troubled. The morphological fact which parallels this psychological performance, and predicts it, is the existence of derivational adjectives based on the verbs of these sentences which are attributive to the subject and the non-existence of adjectives attributive to the object.

The verbs of Table 1(b) reverse the morphological pattern of those in 1(a); i.e., there are derivational adjectives attributive to the sentence object but none attributive to the sentence subject. The causal ratings are also reversed. The majority of experimental subjects gave to the sentences of 1(b) responses of the type 'O >S'. In Table 1, as a whole, then it is as if the derivational adjective based on the verb reflected the direction of greater causal weight (to either the subject or the object) in the semantic of the verb.

When the results of Table 1 (which are representative of our full results) are thought of in terms of grammatical subject and grammatical object, it is possible to say that the greater causal weight is assigned either to subject or object according to whether the derived adjective happens to be attributive to subject or object, but it is not possible to say anything about which it will be, an adjective modifying the subject or an adjective modifying the object, in the case of any particular verb. One must just consult one's intuition or else an English dictionary to find out what adjective has been derived from the verb. So long as the results are conceptualized in terms of subject and object, one cannot detect any consistency in the derivational process itself. In fact, however, the derived forms of Table 1 (and, we will show, of English generally) seem to have been coined according to a semantic principle. Sentence subject and sentence object are, of course, purely syntactic constructs (grammatical relations) which cannot be defined semantically. However, there are in English certain semantic roles (or predicate arguments) coordinated with subject and object, and if we think of the results in Table 1 in terms of semantic roles, a principle governing the creation of derived adjectives will become visible.

Two pairs of semantic roles are associated with subject and object in the class of verbs under consideration: Agent–Patient and Stimulus–Experiencer. Representative definitions (e.g., Chafe, 1970; Comrie, 1981; Fillmore, 1968) of these roles are:

> *Agent.* Someone or something which causes or instigates an action. Usually animate but not always, an agent must be perceived to have its own motivating force.

Patient. Someone or something suffering a change of state.
Stimulus. Someone or something giving rise to a certain experience.
Experiencer. Someone having a given experience.

In English, Agents are almost always grammatical subjects (excepting with the verb *receive*). In the class of verbs that concern us here, all Agents are grammatical subjects and all Patients are objects. Therefore, it is not possible to separate these semantic roles from the coordinated grammatical relations. However, such a separation is possible for Experiencer and Stimulus.

For about half the relevant English verbs the Stimulus role is the grammatical subject and the Experiencer the grammatical object (e.g., *attract*, *deceive*, *charm*), but in the remaining cases, the Experiencer is the subject and the Stimulus the object (e.g., *like*, *loathe*, *detest*). The results of Table 1 can, therefore, be regrouped and relabeled as in Table 2. We now see that, for Stimulus–Experiencer relations, the adjective derived from the verb is always attributive to the Stimulus, whether the Stimulus is sentence subject or sentence object, and, in addition, experimental subjects, in the majority of cases, assign the greater causal weight to the Stimulus. With Agent–Patient relations, the derived adjective is attributive to the Agent and the greater causal weight is, in the majority of cases, assigned to the Agent.

The English verbs associated with Agent–Patient relations belong to a different semantic class than the verbs associated with Stimulus–Experiencer. The first set, which includes *help*, *compete*, *disobey*, *criticize*, *cheat*, *find*, *answer*, etc., may be called 'action verbs'. They name voluntary actions which, in English, are readily used in the imperative mode: *Please help me*; *Please don't compete*; etc. The second set, which includes *charm, attract, like, notice, admire, see, hear*, etc., may be called 'state verbs'. They name involuntary mental states, affective, sensorial or cognitive, which in English resist the imperative mode. It is not unthinkable that one would say: *Please attract me* or *Please don't detest me*, but such sentences are less semantically felicitous than imperatives with action verbs. Dictionary definitions of verbs generally make an action-state distinction. Action verb definitions use forms like *to do*, *to act*, *to express* whereas state verb definitions employ forms like *to feel*, *to be*, *to inspire*. Quite a few English verbs have both action and state senses as, for instance, *mourn*: 'to feel or express great regret'. Dictionary definitions in conjunction with our intuitions about the naturalness of imperative forms serve to identify action and state verbs and verbs having both action and state senses. (See Miller and Johnson-Laird, 1976, for a detailed and rigorous discussion of verb-classes and linguistic membership tests.)

The results of Table 2 are only a regrouping of the results of Table 1, a regrouping in terms of semantic roles and state and action verbs. The re-

grouping suggests a new generalization: In English, adjectives derived from state verbs are attributive to the Stimulus rather than the Experiencer and adjectives derived from action verbs are attributive to the Agent rather than the Patient. This is a generalization that cannot be captured with the grammatical concepts subject and object. The generalization goes far beyond the 12 verbs of Table 2. The evidence for the claim must delayed for a few pages, but what we shall find is that in English generally there are far more derivational dispositional adjectives attributive to the Stimulus than there are adjectives attributive to the Experiencer (about 18 times as many), and there are

Table 2. *Characteristic results formulated in semantic rather than syntactic terms*

(a) Agent–Patient relations; Agent as subject and Patient as object

Sentence	Derived Adjective		Agent > Patient	Patient > Agent	Agent = Patient
Ted helps Paul	*helpful*	[*helpable*?]	15	0	5
Ted cheats Paul	*cheating*	[*cheatable*?]	17	0	3
Ted competes with Paul	*competitive*	[*competitable with*?]	13	1	6
Ted criticizes Paul	*critical*	[*criticizeable*?]	15	0	5

(b) Stimulus–Experiencer relations; Stimulus is subject and Experiencer is object

			S > E	E > S	S = E
Ted attracts Paul	*attractive*	[*attractable*?]	15	2	3
Ted charms Paul	*charming*	[*charmable*?]	18	1	1
Ted deceives Paul	*deceptive*	[*deceivable*?]	15	1	4
Ted troubles Paul	*troublesome*	[*troublable*?]	15	0	5

(c) Stimulus–Experiencer relations; Experiencer is subject and Stimulus is object

			S > E	E > S	S = E
Ted likes Paul	*likable*	[*liker*; *-ing*; *-ful*?]	13	1	6
Ted detests Paul	*detestable*	[*detester*; *-ing*; *-ful*?]	16	1	3
Ted loathes Paul	*loathsome*	[*loather*; *-ing*; *-ful*?]	17	2	1
Ted notices Paul	*noticeable*	[*noticer*; *-ing*; *-ful*?]	15	1	4

far more derivational dispositional adjectives attributive to the Agent than there are adjectives attributive to the Patient (about 14 times as many). These derivational processes appear to operate in English (perhaps also in other languages) in an extremely biased, asymmetrical fashion. It is as if speakers of English tend to locate the primary dispositional causes of actions in Agents and of states in Stimuli. This is a very strong claim which the reader can only be asked to entertain, at this point, for the purposes of the general argument.

If we think, prematurely but briefly, of models for the process of assigning causal weights in the standard task, the generalizations suggested by Tables 1 and 2 admit of two quite different possibilities. One possibility is a kind of priming process, in a network that includes both the verb and the derivational adjective. It may be that when a particular verb (e.g., *help*) is activated by presentation of a sentence (e.g., *Ted helps Paul*), that activation flows to any adjective derived from the verb stored in long-term memory (e.g., *helpful*) and that the definition of the adjective (e.g., 'the kind of person that helps people') stored with the adjective mediates the choice between rating more highly 'Alternative A' or 'Alternative B'. This chain may be represented as: verb → adjective → definition → causal weight. The priming model would work perfectly well if there were no principle governing the creation of derived adjectives in English. It would simply mean that wherever English happened, for whatever accidental reasons, to have laid down a dispositional derived adjective, the preponderant causal weight would be assigned to that argument of the verb that was modified by the adjective.

The fact that there does seem to be a principle governing English derivations, without in any way militating against the priming hypothesis, opens the way to a quite different sort of processing model. Perhaps what happens when a subject encounters a verb in a sentence (e.g., *help* in *Ted helps Paul*) together with a causal rating to be made is that the perception of the verb accesses a lexical entry in memory which includes syntactic class (verb); subclass (action or state); associated semantic roles or verb arguments (Agent–Patient or Stimulus–Experiencer); grammatical subject and object assignment of the roles; and relative causal weights; with redundancies eliminated. In the case of *help*, the corresponding entry would specify: verb; action; Agent—Patient; Agent is subject; Agent is primary cause. From this sort of information subjects could make their causal ratings without in any way consulting the derived adjective (*helpful*). For this kind of processing model, let us call it a causal schema model, the existence or not of derived adjectives based on the verb is of no consequence. The processing does not in any way involve the particular derived form.

As far as the two processing models are concerned, the existence of a

principle governing the creation of derived forms introduces a kind of causal confound. Do the ratings subjects make depend on particular mediating adjectives or only on the schemas suggested by the principle governing the creation of such adjectives? There is some possibility of separating the confounded factors. Not every derived adjective in English conforms to the general principle of derivation: *commendable* names a Patient disposition and *appreciative* an Experiencer disposition. Would subjects given the sentence *Ted commends Paul* assign the greater weight to Ted in accordance with the Agent–Patient Schema or to Paul in accordance with the existent adjective and the priming hypothesis? With the sentence *Ted appreciates Paul*, would it be *Paul* in accordance with the Stimulus–Experiencer Schema or *Ted* in accordance with *appreciative* and the priming hypothesis? Research designed to choose between the two sorts of processing models is in progress but will not be reported in the present paper.

Whichever kind of processing model is involved, priming or schema, the results of Tables 1 and 2 would seem to be a kind of Whorfian result, an effect of language (either of particular derived adjectives or of a general principle of derivation) on thought (the distribution of causal weights). Our own guess is that it is not a Whorfian result. There is, at present, just one strong reason for thinking not. The English language, viewed as a system of possible constructions (*langue* not *parole*) is not biased toward either party in the interactions named by verbs like *help*, *disobey*, *charm*, *attract*, *delight*, *like*, *value*, and *detest*. English derivational morphology (Francis, 1958; Aronoff, 1976) provides some suffixes for transforming verbs to adjectives which are restricted to particular verbs or verb classes:

-ful	*helpful, harmful, delightful, scornful*
-iant; ient	*disobedient, defiant, compliant*
-ive	*attractive, repulsive, protective*
-some	*bothersome, loathsome*
-ious	*censorious, injurious*
-worthy	*praiseworthy, blameworthy, trustworthy*

And so on. In addition, however, there are largely unrestricted fully productive suffixes which can be used to create dispositional forms either for any Agent or for any Patient, for any Stimulus or any Experiencer. These are *-er* and *-ing* for sentence subjects and *-able* (*-ible*, etc.) for sentence objects. We are very familar with *reader*, *speaker*, *teacher*, have come to know *computer-programmer*, and would not balk at *viewer-with-alarm*. We are very familiar with *charming*, *comforting*, *shocking*, have had no problem with *caring* (person), *castrating* (person), and would understand, if we heard them, *valuing* (person) or *detesting* (person). We know *likable*, *valuable*, *detestable*, and

when Alexander Haig described himself as *unshockable*, we knew at once what he meant. A television *policier* entitled 'Muggable Mary' concerned a female officer got up to attract muggers, and so the new coinage *muggable* was apt and in no need of definition.

All of the derived forms marked with a question mark in Table 2 are, in fact, *possible* English words, though perhaps not *actual* English words. *Helpable*, *charmable*, *attractable* are, all of them, perfectly understandable. A *liking* person or a *liker* are just as possible as a *loving* (person) and a *lover*. The fact that English morphology makes available such generally productive forms as *-er*, *-ing*, and *-able* means that English *as an abstract generative system* does not exclude the questioned forms of Table 2. Interesting effects of language on thought must refer to language as a system not to particular speech acts if only for the reason that there can never have been any doubt that speech acts exert an effect on thought.

Can it be said that the lexicon of English, as a list of *actual* words, does exclude the questioned forms? Perhaps, but it is not easy to locate that official Ultimate Lexicon. The *Oxford English Dictionary* (1933) includes independent adjective entries for many derived forms most of us would not recognize as actual; for instance, Haig's *shockable* as well as *amazable*, *amusable*, and *hateable*. *Webster's Third Edition Unabridged* (1961) and the *Random House Unabridged* (1966) exclude some of these as obsolete but still contain many that would not be recognized by everyone; for instance, *agitable*, *depressible*, and *fightable*. In fact, no firm line can be drawn between actual words and possible words. The distinction is a fuzzy one. The opening sentence of the *Oxford English Dictionary* says it once and forevermore: "That vast aggregate of words and phrases which constitutes the Vocabulary of English-speaking men presents, to the mind that endeavours to grasp it as a definite whole, the aspect of one of those nebulous masses familiar to the astronomer, in which a clear and unmistakeable nucleus shades off on all sides, through zones of decreasing brightness to a dim marginal film that seems to end nowhere, but to lose itself imperceptibly in the surrounding darkness".

For our preliminary work we have drawn a line between actual and possible words by using a criterion that is arbitrary-but-objective. A derived form is considered to be actual if it appears with a frequency of one-in-one-million or greater in the Kučera–Francis *Computational Analysis of Present-Day American English* (1967). By this criterion *helpful*, *cheater*, *competitive*, *critical*, *attractive*, *charming*, *deceitful*, *troublesome*, *likable*, *detestable*, *loathesome*, and *noticeable* are all actual words whereas their questioned counterparts in Table 2 are not. Since it is a frequency criterion that defines our actual derived forms, it is not the English language as a generative system that favors *helpful* over *helpable*, *likable* over *liking*, but rather the usage

of English speakers, and it is also not the English language that is biased toward the creation of derived forms for Agents rather than Patients, Stimuli rather than Experiencers; it is the habits of speakers. What are we to make of this?

Our hypothesis is that humans, speakers of whatever language, operate with two basic causal schemas in the domain of psychology and that the way we talk is a consequence of these schemas, not the cause of them. For actions, that is behavior involving the voluntary musculature, we attribute causality primarily to dispositions in the Agent conceived to be relatively general across Patients. Thus, an instance of helping seems to us to arise out of a disposition to be helpful—quite generally across Patients—rather than a tendency to instigate or merit help in certain persons. This way of thinking, the Agent–Patient Schema, results in the creation of many more Agent dispositional terms than Patient dispositional terms. For mental states, especially perhaps affective states of the autonomic nervous system, we attribute causality primarily to dispositions in the Stimulus conceived to be relatively general across Experiencers. Thus, an instance of someone being charmed seems to us to arise out of a disposition of a Stimulus to charm, to be charming—quite general across Experiencers—rather than from a tendency to be easily charmed in certain persons. This way of thinking, the Stimulus–Experiencer Schema, results in the creation of many more Stimulus dispositional terms than Experiencer dispositional terms. Readers familiar with basic distinctions in learning theory will notice that the Agent-Patient Schema identifies responses of the type that B.F. Skinner (1938) calles 'operants'—responses an organism *emits*—whereas the Stimulus-Experiencer Schema identifies responses of the type B.F. Skinner calls 'respondents'—responses that are *elicited* by stimuli.

If the Agent-Patient Schema and the Stimulus-Experiencer Schema represent human ways of thinking about psychological causality, it follows that in all languages, most importantly in languages not historically related to English, Agent dispositions should, on the whole, be more 'codable' (Brown and Lenneberg, 1954; Brown, 1958) than Patient dispositions and Stimulus dispositions more codable than Experiencer dispositions. For English the difference of codability comes to the difference between an actual word (e.g., *charming*) and either a possible word (e.g., *charmable*) or a phrase (e.g., *easy to charm*). In some closely related languages, the difference is of the same type; in French *charmant*, *charmable*?, or *facile à charmer*. In unrelated languages one does not, of course, expect to find derivational morphology of a kind comparable to English, and so it remains to be determined just how differences of codability will be manifested. Work is now underway investigating the differential codability hypothesis for Chinese and Japanese, but is not yet ready to be reported. If the schemas are found to operate in such

unrelated languages, the non-Whorfian, thought affects language hypothesis, will be strongly supported. The argument set forth here based on the fact that frequency of usage and not rules of morphology distinguished actual derived adjectives from possible derived adjectives for English can only serve as a first clue.

Evidence supporting the argument

First we shall have the evidence that in the derivation of adjectives from verbs English has, for state verbs, favored adjectives attributive to the Stimulus over adjectives attributive to the Experiencer and, for action verbs, has favored adjectives attributive to the Agent over adjectives attributive to the Patient. Then, for a randomly selected sample of 36 verbs, comes evidence that the greater causal weight is assigned the Stimulus rather than the Experiencer, the Agent rather than the Patient. This evident was obtained using the standard task describing the Overview with sentences in the active voice and also with sentences in the Passive voice. Finally, evidence is presented that the differential causal weights are obtained with formats very different from the standard.

A study of the lexicon

The method is best described in sequential steps.

1. In alphabetical order we went through the entire Kučera-Francis list of English words that occur one-or-more times in one million words of text. We extracted all dispositional adjectives based on verbs, such as *abhorrent*, *abusive*, *accommodating*, etc.
2. For the dispositional adjective to be retained on our list, it was necessary that the verb base meet the following test:
   semantic and syntactic acceptability in the contexts

$$\textit{Ted} \\ \textit{Mary} \end{Bmatrix} - \begin{Bmatrix} \textit{Paul} \\ \textit{Ann.} \end{Bmatrix}$$

   The only reason for using both male and female proper names is the slight awkwardness of some verbs (e.g., *adore, dazzle*) on same gender pairings.
3. For the dispositional adjective to be retained on our list, it had to be acceptable in a context (following the verb context above):

$$\textit{Why? Because} \begin{Bmatrix} \textit{Ted} \\ \textit{Mary is} \text{ —— .} \\ \textit{Paul} \\ \textit{Ann} \end{Bmatrix}$$

   Thus *Mary charms Paul. Why? Because Mary is charming.*

The form *charming* went on the list. Notice that the second context excludes derived forms using the suffix *-ed*. Thus, *Mary charms Paul. Why? Because Paul is charmed.** The *-ed* form is a derived form, attributive here to the Experiencer, but it is not a derived form naming a disposition that can count as a cause of the interaction depicted by the sentence whereas *-ing* forms often do name such dispositions. Paul's state of being charmed (resulting from his interaction with Mary) can help to cause subsequent actions or states of Paul but not the state named in the sentence. The fact that there is an actual word for the disposition created by Mary's charm should not have anything to do with causal weights in the creation of Paul's charmèd state. Our thesis is not that there are no (or few) derived forms naming any sort of Experiencer (or Patient) disposition, but that there are few relevant to the causality of the interaction named by the verb base.

4. Derived forms with the suffix *-ing* (e.g., *charming*) present a special problem. The Kučera–Francis count does not distinguish between *charming* as a verb participle and *charming* as an adjective, nor does any other published frequency count, and so one cannot tell how much of the listed frequency is verb usage nor even whether the presence of the form on the list at all results from its use as an adjective. From Professor W. Nelson Francis of Brown University we have learned that a new grammatically-tagged frequency list that counts separately adjectival and verbal *-ing* is in preparation. Not having just the kind of count needed, we used the following procedure. A form with *-ing*, in order to be counted as a dispositional term, had to have a main entry as an adjective in *Webster's Third International Unabridged Dictionary*. *Amusing* and *charming* have such main entries whereas *antagonizing*, *arousing*, and *solacing* do not but are simply listed under the respective verbs as inflected forms. The distinction seems to have been based (roughly) on frequency in an adjectival sense.
5. To make an independent check on the dispositional status of the derived forms on our list, we looked for them on the Allport-Odbert "List of Terms in the English Language Characterizing Personal Behavior and Personality". The Allport-Odbert list "contains all the words descriptive of personality or personal behavior (save those that are obsolete) included in *Webster's New International Dictionary*" (1936, p. 24). The edition of Webster's used (1925) "comprises approximately 400,000 separate terms or derivatives ... The criterion for inclusion consists in the capacity of any term to distinguish the behavior of one human being from that of another. Terms representing common (non-distinctive) behavior are excluded, e.g., *walking* and *digesting* ..." (1936, p. 24). The

Allport–Odbert List contains 17,953 words. In the process we have come to know very well this fascinating work, which Gordon Allport and Henry Odbert characterized as a 'psycho-linguistic' study; the present-day absence of the hyphen is a good psycholinguistic index of the maturity of our science.

6. It was finally necessary to identify the verb bases of the dispositional terms on our list as actions or states or both. (We recognize that our treatment of verbs may be an oversimplification since some linguists believe that all verbs have, in varying degree, both active and stative possibilities.) We used two criteria. Definitions of verbs in *Webster's Third Unabridged* were consulted for indications (usually present) as to whether the verb is an action (*to do*, *to express*, etc.) or a state (*to feel*, *to be affected by*, etc.). We consulted our own intuitions as to the naturalness of the verb in the imperative mode; the less natural forms are state verbs. Where a verb has both state and action senses (e.g., *encourage*, "to inspire confidence or express approval"), we counted it as both. Thus, in the action sense of *encourage* (*Ted encourages Paul*), *Ted* is Agent and *Paul* is Patient whereas, in the state sense, *Ted* is Stimulus and *Paul* is Experiencer.

Several linguists and philologists have studied our procedures and commented on them to us. From their comments we have learned that morphological arguments can be made for one or another slight change but that the tests used are reasonably apt and straightforward. They left us with a list of dispositional adjectives derived from state verbs and a list of dispositional adjectives derived from action verbs. The hypothesis was that for state verbs there would be more adjectives attributive to the Stimulus than adjectives attributive to the Experiencer and that for action verbs there would be more adjectives attributive to the Agent than adjectives attributive to the Patient. The results are:

State verbs	
Dispositional forms for the *Stimulus* only	144
Dispositional forms for the *Experiencer* only	8
Dispositional forms for both *Stimulus* and *Experiencer*	18
Action Verbs	
Dispositional forms for the *Agent* only	138
Dispositional forms for the *Patient* only	10
Dispositional forms for both *Agent* and *Patient*	13

An example of a verb from which only an Agent disposition is derived is *defy* → *defiant*. A Patient Only verb is *avoid* → *unavoidable* (note that negative forms were included if they met other criteria). An Agent and Patient

verb is *corrupt* → *corrupter* and *corruptible*. A Stimulus Only verb is *despise* → *despicable*. An Experiencer Only verb is *disdain* → *disdainful*. A Stimulus and Experiencer verb is *trust* → *trustful* and *trustworthy*. When there were multiple derived forms of the same dispositional type for a given verb (e.g., *deceitful* and *deceptive*; *neglectful* and *negligent*), we made only one tally; this was a conservative decision since most multiple forms were of types favoring the hypothesis.

The simple frequencies listed above show that the hypotheses were confirmed. There were about 14 times as many Agent Only cases as Patient Only and 18 times as many Stimulus Only as Experiencer Only. Of all Agent or Patient attributions, 87% were Agent, and of all Stimulus or Experiencer attributions, 86% were Stimulus.

A number of interesting things turned up in this study, and of these we will mention only one. There were quite a few dispositional forms carrying the suffix *-ful*. Concerning this suffix, the *Oxford English Dictionary* says that it was originally identical with *full* and when used in the composition of adjectives has the sense 'full of ...' as in *scornful* or *disdainful*. It is interesting, therefore, to find on our list the following: *delightful*, *wonderful*, *awful*, *dreadful*, *shameful*, *hateful*, *frightful*, *fearful*, *painful*, *pitiful*. All of these (according to the *O.E.D.*) had the 'full of' sense originally and so would have been Experiencer dispositions (as *scornful* and *disdainful* are). For example:

> I am so frightful at being in a murderer's house (1802);
> The Turks, of whom the ladies take a dreadful view (1659).

Today, however, either the only, or the most common, sense of *-ful* forms is a Stimulus disposition. Someone who is *delightful*, *wonderful*, or *dreadful* is not full of the feeling named but is likely to elicit the feeling named. Where both senses are still alive, we have used the *Random House Dictionary* as our criterion for the more common sense today since the editors say that they have ordered their definitions by this criterion.

Certainly we need the advice of philologists on this point, but before we get to them and they (inevitably) shoot us down, it is fun to suggest that the basic Stimulus–Experiencer schema has brought these forms into line, changing them from Experiencer dispositions to Stimulus dispositions. The only example we have found of a semantic change in the opposite direction is *trustful*, which was not coined until the early 19th century (*O.E.D.*) and at first meant inspiring trust, which is today the sense of *trustworthy*.

Causal ratings of three classes of verbs

This is the 'Ted likes Paul' paper-and-pencil task. The directions are an expanded version of those quoted in the Overview. Experimental subjects were

20 Harvard-Radcliffe undergraduates. The verb set consists of 36 verbs divided into three classes of 12 verbs each. Class I verbs are Agent–Patient forms with the Agent as sentence subject (*apologize*, *cheat*, *criticize*, etc.). Class II verbs are Stimulus–Experiencer forms with the Stimulus as sentence subject (*astonish*, *attract*, *charm*, etc.). Class III verbs are Experiencer–Stimulus forms with the Experiencer as sentence subject (*abhor*, *admire*, *despise*, etc.). The 36 verbs are listed in Table 3.

This experiment was done before the study of the lexicon described above and, indeed, it was the very striking results of this experiment that motivated the much more time-consuming study of the lexicon. Because of the way the studies were ordered in time, the 36 verbs could not be taken from a final complete list as that list was not yet in existence. Therefore, the verbs were selected by entering the Kučera–Francis alphabetical list at 36 random points and moving through the list until finding a derived form of a targeted type with frequency greater than one in one million.

Table 3 reports the results for all verbs of all three classes. The results are uniformly reported in terms of sentence-subject (S) and sentence-object (O), and so it is necessary to note that our prediction is S > O for Class I and Class II in which sentence-subjects are, respectively, Agent and Stimulus. For Class III verbs the prediction is O > S since the sentence-subject is, in this case, the Experiencer. In Table 3 the difference in the mean ratings of subject and object is in the predicted direction for every single verb. Thirty-two of 36 mean differences are significant with $p \leq 0.05$, and 16 are significant with $p \leq 0.001$. These verb ratings operate almost like test items with factual answers and yet the knowledge involved is completely out of awareness. We have always asked S's to try to guess what the hypotheses underlying the questionnaire might be, and none has ever come close.

Not knowing what kind of measurement scale the 9-point ratings may constitute, results are reported in Table 3 in a form that assumes only nominal properties. An experimental subject's ratings for any particular verb may be classified as S > O, O > S, or S = O. This three-category scoring can be reduced to a dichotomy by assuming that subjects responding S = O would, if a choice were forced, divide equally often into S > O and O > S and so distributing S = O tallies evenly between S > O and O > S. The results can then be tested very simply for significance using the binomial distribution and as null hypothesis S > O = O > S = 0.5. A one-tailed test (the direction of the difference was predicted) with $n = 20$ requires that the smaller entry be six or fewer for a p of about 0.05; four or fewer for a p of about 0.01; and three or fewer for a p of 0.001 or better. As in the case of the means the differences are in the predicted direction for all verbs. Of 36 verbs, differences are significant at 0.05 or better for 28.

Table 3. *Causal weightings for three classes of verbs: Active voice* ($n = 20$)

Class I. Agent–Patient Subject–Object	Subject mean	Object mean	S > O	O > S
apologize-to	5.55	3.2 **	16.0	4.0**
cheat	7 7	4.4 ***	18.5	1.5***
compete-with	7 7	5.35***	16.0	4.0**
criticize	6.95	4.5 ***	17.5	2.5***
defy	6.2	4.0 **	16.0	4.0**
disobey	5.6	5.15	12.0	8.0
dominate	7 55	6.7	11.5	8.5
flatter	7.4	6.1 **	14.0	6.0*
harm	6.7	3.4 ***	19.0	1.0***
help	7.6	5.15***	17.5	2.5***
protect	6.65	5.45*	13.5	6.5
slander	6.9	3.75***	17.5	2.5***
Class II. Stimulus–Experiencer Subject–Object	**Subject mean**	**Object mean**	**S > O**	**O > S**
astonish	6.65	3.8 ***	18.5	1.5***
attract	6.85	3.95***	16.5	3.5**
charm	7.2	3.25***	18.5	1.5***
deceive	7.15	4.4 ***	17.0	3.0***
delight	6.2	3.9 **	17.0	3.0***
exasperate	6.5	4.4 **	16.0	4.0**
impress	7.3	4.1 ***	19.5	0.5***
influence	7.15	5.8 *	14.5	5.5*
repel	5.7	5.05	12.5	7.5
scorn	6.05	4.7 *	13.5	6.5
shock	6.2	4.45*	16.5	3.5**
trouble	6.2	3.95**	17.5	2.5***
Class III. Experiencer–Stimulus Subject–Object	**Subject mean**	**Object Mean**	**S > O**	**O > S**
abhor	4.05	5.5 *	8.5	11.5
admire	4.1	8.05***	1.0	19.0***
despise	4.2	6.05**	5.0	15.0
detest	3.2	6.15***	2.5	17.5***
dread	3.8	6.15**	5.5	14.5*
enjoy	5.8	6.75*	6.5	13.5
esteem	4.2	7.2 ***	1.0	19.0***
honor	4.05	7.0 ***	1.5	18.5***
like	5.7	7.05**	4.0	16.0**
loathe	3.8	6.25***	2.5	17.5***
notice	5.3	7.2 ***	3.0	17.0***
pity	5.15	6.25	5.5	14.5*

S > O or O > S: $p \leq 0.05^{*}$, $p \leq 0.01^{**}$, $p \leq 0.001^{***}$.

Causal ratings with sentences in the passive voice

Instead of *Ted likes Paul*, the depicting sentence becomes: *Paul is liked by Ted*. The alternatives are as before: A. 'Ted is the kind of person that likes people'; B. 'Paul is the kind of person that people like'. The first causal rating study (36 verbs and 20 subjects) was repeated with all sentences in the passive voice to check the possibility that the results originally obtained depended on something other than the meanings of the verbs and their related derived forms; something like the surface subject and object of the sentence. The results appear in Table 4 in a form paralleling Table 3 (causal ratings for active-voice sentences). Sentence subject and object refer to deep or logical subject and object and so are the same in Table 4 as in Table 3.

In Table 4 the differences in the means are in the predicted direction in all 36 verbs and so are the differences between S > O and O > S. Of the 36 mean differences, 33 are significant with $p \leqslant 0.001$. The tallies yield a similar picture. Causal ratings made for sentences in the passive voice produce results which, like the ratings for sentences in the active voice, almost suggest test questions with factual answers.

Within class rank-order coefficients for the means between active voice and passive voice were: Class I, rho = 0.76; Class II, rho = 0.83; Class III, rho = 0.69 All three correlations are significant with $p \leqslant 0.01$. In short, the individual verbs tend to 'breed true' to convey the same causal balance in the passive voice as in the active voice. Clearly the surface arrangement of sentence subject and object does not affect causal ratings. These ratings seem to depend on the meanings of the verbs. These meanings fall into the two major classes described by the Agent–Patient Schema and the Stimulus–Experiencer Schema. There is, however, some residual intra-class variation in relative causal weights and this variation seems to be a stable property of the particular verbs.

Free listing of causes

The standard task which asks whether 'Ted is the kind of person that ...' or 'Paul is the kind of person that ...' is quite closely connected to the derived adjectives since the alternatives offered would be acceptable as definitions of the adjectives; *likable* means the kind of person people like, and *helpful* means the kind of person that helps people. Are the differential causal ratings obtained for verbs of Classes I, II, and III obtainable only with such a fairly directive format? To answer that question we devised a new questionnaire:

> Each item in this booklet consists of a short sentence depicting a simple psychological event involving two persons, A and B, followed by the question

Table 4. *Causal weightings for verbs in the passive voice* ($n = 20$)

Class I.	Agent–Patient Subject–Object	Subject mean	Object mean	S > O	O > S
apologize-to		6.65	3.4 ***	19.0	1.0***
cheat		7.55	4.2 ***	19.0	1.0***
compete-with		6.8	3.95***	17.0	3.0***
criticize		7.55	5.25***	17.0	3.0***
defy		6.55	5.25*	13.5	6.5
disobey		7.0	5.1 ***	16.5	3.5**
dominate		8.05	6.8 **	15.0	5.0*
flatter		7.4	5.0 ***	17.0	3.0***
harm		6.7	4.15***	17.0	3.0***
help		7.6	4.9 ***	16.5	3.5**
protect		6.3	6.15	10.5	9.5
slander		7.8	3.8 ***	19.0	1.0**
Class II.	Stimulus–Experiencer Subject–Object	Subject mean	Object mean	S > O	O > S
astonish		6.6	4.25***	16.0	4.0
attract		7.15	3.7 ***	17.5	2.5***
charm		7.25	4.0 ***	19.5	0.5***
deceive		7.65	4.5 ***	18.5	1.5***
delight		6.95	4.75***	14.5	5.5*
exasperate		6.9	4.95**	17.0	3.0***
impress		7.5	5.3 ***	16.0	4.0**
influence		7.35	6.5	12.0	8.0
repel		6.75	4.9 **	14.0	6.0*
scorn		6.7	5.75*	13.0	7.0
shock		6.65	5.3 *	13.5	6.5
trouble		6.15	4.9 *	13.5	6.5
Class III.	Experiencer–Stimulus Subject–Object	Subject mean	Object Mean	S > O	O > S
abhor		5.1	6.7 **	5.5	14.5*
admire		4.65	7.7 ***	1.5	18.5***
despise		5.75	6.45	8.0	12.0
detest		5.3	6.8 *	6.0	14.0*
dread		5.5	7.7 ***	2.0	18.0***
enjoy		6.4	7.45*	7.0	13.0
esteem		4.25	7.55***	3.0	17.0***
honor		4.85	7.1 ***	3.0	17.0***
like		5.8	6.9 *	5.5	14.5*
loathe		4.95	6.4 *	5.0	15.0*
notice		6.05	7.85***	1.5	18.5***
pity		5.6	7.2 ***	3.5	16.5**

S > O or O > S: $p \leq 0.05$*, $p \leq 0.01$**, $p \leq 0.001$***.

> 'Why?' You are to answer the question in a short sentence beginning: 'Because ...' and continuing by saying something about A or else something about B. For instance, the sentence might be: 'A loves B'. In response to the question 'Why?' you might write: 'Because A is a warm-hearted person' or, perhaps, 'Because B is very beautiful'. You must, however, not say something about both A and B, but only something about one or the other. It will always be possible, of course, to give a reason for the event in terms of the qualities of either A or B. What you are asked to do is to choose between them.

Subjects were 20 Harvard–Radcliffe students. Asked why A likes B, subjects wrote such things as: 'Because B is kind'; 'Because B has a warm crooked grin'; 'Because B is generous'. Asked why A helps B, they wrote such things as: 'Because A needs money'; 'Because A wants what B has'; 'Because A is dishonest'. Very rarely did anyone write down the relevant derived adjective (e.g., *likable* or *helpful*); probably the directions lead subjects to suppose that something less tautological was expected. There were also not many duplications in the set of causes listed; most were unique. However, subjects were like one another in one respect, the critical respect: they mostly answered questions involving Class I verbs with sentences beginning 'Because A ...' (the Agent and subject) and sentences with Class II verbs with 'Because A ...' (the Stimulus and subject), but sentences with Class III verbs began 'Because B ...' (the Stimulus and object).

The results appear in Table 5. Of 36 verbs, the differences are in the predicted direction for 31. For the constrained causal weightings task (Tables 3 and 4) differences were in the predicted direction in all 36 cases. Twenty-four of 36 differences are significant with $p \leq 0.05$ (28 in Table 3 and 27 in Table 4) and 12 with $p \leq 0.001$ (17 in Table 3 and 15 in Table 4). In short, the method of free listing of causes yields very strong results for verbs of Class I, Class II, and Class III consistent with the two schemas and with the derived adjectives of English. The results are slightly less strong than those obtained with the more constrained format and that is presumably because the open task affords more apportunity for incidental unsystematic associations to operate. Considering how great the opportunity is for such associations, it seems remarkable that the results are, overall, so very systematic.

With Table 5 the time has come to acknowledge that one verb in our standard set of 36 is misclassified. The misclassification was noticed at a late point in our work by a very careful reader and we decided to let it stand because the unintentional mistake might prove instructive. The verb *scorn* is not a Class II verb with the Stimulus as subject and Experiencer as object. *Scorn* in its stative sense (to feel scorn) has as sentence subject the Experiencer; in *Ted scorns Paul* it is Ted that feels scorn and Paul that inspires the feeling. This fact alone would make *scorn* a Class III verb (Experiencer as

Table 5. *Free listing of causes for three classes of verbs* (n = 20)

Class I	[Agent–Patient [Subject–Object	A	B	
apologize to		17	3	***
cheat		17	3	***
compete with		15	5	*
criticize		10	10	
defy		12	8	
disobey		8	12	
dominate		11	9	
flatter		15	5	*
harm		15	5	*
help		11	9	
protect		10	10	
slander		11	9	
Class II	**[Stimulus–Experiencer [Subject–Object**	**A**	**B**	
astonish		10	10	
attract		20	0	***
charm		18	2	***
deceive		18	2	***
delight		15	5	*
exasperate		18	2	***
impress		17	3	***
influence		12	8	
repel		12	8	
scorn[+]		5	15	*
shock		13	7	
trouble		15	5	*
Class III	**[Experiencer–Stimulus [Subject–Object**	**A**	**B**	
abhor		4.5	15.5*	
admire		2	18	***
despise		4	16	**
detest		4	16	**
dread		3	17	***
enjoy		4	16	**
esteem		3	17	***
honor		3	17	***
like		3	17	***
loathe		5	15	*
notice		1	19	***
pity		3	17	***

A > B or B > A: $p = \leq 0.05$*, $p \leq 0.01$**, $p \leq 0.001$***.
[+] Misclassified.

subject). However, the only English adjective derived from *scorn* is *scornful* (*scornable* is absent) and *scornful* is attributive to the Experiencer. The verb *scorn*, therefore, is used derivationally in a way that violates the Stimulus–Experiencer Schema. It is one of the eight such violations we have identified in English and, as such, does not belong with the other 35 verbs in our set since for all these the derived forms follow the rule of the two schemas.

Scorn is, in fact, the sort of verb to make possible a test of strength between the two normally confounded processing determinants: the mediating adjective primed by a verb and the causal schema. The adjective *scornful* predicts a greater causal weight (in 'A *scorns* B') for 'A'—the Experiencer—whereas the Stimulus–Experiencer Schema predicts a greater causal weight for 'B'—the Stimulus. In Table 5 the schema rather strikingly prevails over the adjective; the main causal weight was assigned to 'B', the Stimulus, which makes *scorn* look out of place in Class II (Stimulus is 'A') whereas it would fit nicely into Class III (Stimulus is 'B'). Score one point, then, for the schema model over the priming model. If, however, we look at the entries for *scorn* in Tables 3 and 4, it becomes evident that a final judgment is not so easily attained.

In Tables 3 and 4 *scorn* does not look out of place in Class II and would look out of place in Class III for the reason that greater causal weight is assigned the subject (the *scornful* Experiencer) than the object. To be sure, we see that *scorn* is among the verbs in Class II that show relatively small differences between subject and object, and this fact together with the reversal of Table 5 may perhaps mean that in an anomalous verb like *scorn* both schema and derived adjective exert effects, opposite in direction, on the causal balance. However, even this vague conclusion is not clearly justified. The problem is that *scorn* has not only a stative sense but also an action sense, as in 'Mary scorned (actively rejected) Paul'. The stative sense is probably the stronger of the two, but when the action sense operates, as it may for some subjects, then the Agent–Patient Schema applies and the schema prediction favors the subject, and in Tables 3 and 4 but not 5, the subject is assigned the greater weight. However, if the Agent–Patient Schema applies, then the verb is not derivationally deviant because *scornful* is attributive to the Agent. The reader is now in a position to appreciate the fact that it will not be easy to devise critical tests to choose between priming and schema processing models.

The results with free listing of causes might seem to tell against a priming process model, but, in fact, they do not. They do tell against a model that supposes activation to move from the presented verb (e.g., *help*) to the stored derivational adjective (*helpful*) to the adjective's stored definition ('kind of person that helps') to the corresponding rating alternative ('kind of person

who helps'). The last step is not called for by the instructions. However, it is not unreasonable to suppose that the first three steps occur after which activation flows to whatever word or phrase is most strongly linked in the memory of the individual subject with 'kind of person who helps': *altruistic* or *kind* or *smart*. These final entries should vary with the experience of the individual and tend to be idiosyncratic—as they are. While the results of Table 5 do not exclude a priming process, they also do not exclude a process that operates with general schemas and without particular adjectives. One need only suppose that activation of the Stimulus–Experiencer Schema by a state verb makes causal thinking flow most naturally from the Stimulus, and activation of the Agent–Patient Schema by an action verb produces causal thinking in terms of the Agent.

The calculus of causality

In this experiment subjects were asked to operate with the rules of causal inference as formulated by Kelly (1967). There is a great deal of evidence (e.g., McArthur, 1972; Orvis *et al.*, 1975) that all adults have tacit or implicit knowledge of these rules. The rules map the event to be explained into:

Person Action or State Stimulus

The variable called 'Consensus' concerns generalization across Persons and is usually conceived dichotomously as High or Low. The variable called 'Distinctiveness' concerns generalization across Stimuli and is also either High or Low. The variable called 'Consistency', also just High or Low, concerns generalization across time and/or modality. Particular combinations of values on these three dichotomous variables regularly lead to attributions of particular kinds. Most notably, High values on Consensus, Distinctiveness, and Consistency lead to Person attributions, and Low Consensus with Low Distinctiveness and High Consistency leads to Stimulus attributions. We cannot here take the space for a full exposition of this inferential calculus but must refer the reader to Kelley (1967) and, especially, McArthur (1972). However, we will attempt to convey the gist of the inference process just with respect to Consensus and Distinctiveness. Consistency is not important in our problems and so can be set aside.

We begin with a homely example. Suppose Mr. B has trouble starting his car and that we are interested in locating the cause of this trouble. Initially, we simply want to know whether the problem is in Mr. B or in his car. Let us now map the problem into the concepts of attribution theory:

Person	Action	Stimulus
Mr. B	has trouble starting	his car

Holding the Person and Stimulus constant, we ask the Consensus question: "Do many or few drivers other than Mr. B have trouble starting his car?" If many then the Consensus is High, which means there is a high generalization across Persons, and if few, then the Consensus is Low, which means little generalization across Persons. Holding the Person and Action constant, we ask the Distinctiveness question: "Does Mr. B have trouble starting many or few other cars?" If many, then Distinctiveness is Low, which means high generalization across cars, and if few, then Distinctiveness is High, which means little generalization across cars. Note the way in which the terminology for Consensus and Distinctiveness reverses (as it can be a source of confusion): for Consensus the value High means much generalization and Low, little, but for Distinctiveness High means little generalization (the Stimulus is distinctive) and Low means much generalization (the Stimulus is not distinctive).

There are two patterns of information which regularly lead to particular forms of attribution:

Attribution to:	Person	Attribution to:	Stimulus
Consensus:	Low	Consensus:	High
Distinctiveness:	Low	Distinctiveness:	High

If you test the two patterns on your intuition, they should 'feel' right. The first pattern (hereafter 'LL' for Low Consensus and Low Distinctiveness) says that few drivers other than Mr. B have trouble starting his car and Mr. B has trouble starting many cars other than his own. In this case, we should attribute the trouble to Mr. B, and that is a Person attribution. If the attribution (and not simply its locus) were made specific, we might say something like: 'Mr. B is a poor driver'. The second pattern (hereafter 'HH' for High Consensus and High Distinctiveness) says that many drivers other than Mr. B have trouble starting his car and Mr. B has trouble starting few other cars (his car is distinctive). In this case we should attribute the trouble to Mr. B's car and that is a Stimulus attribution. If the attribution (and not simply its locus) were to be made specific, we might say something like: "This car is hard to start".

It remains to translate our interpersonal interaction sentences (e.g., *Ted likes Paul*) into the concepts of attribution theory:

Person	State	Stimulus
Ted	likes	Paul

Some of the names given to the concepts in attribution theory now prove awkward. There is no problem with the use of 'Action' or 'State', as the case may be, to name the interaction between two entities that is to be explained.

The problem lies in dubbing one of these a 'Person' and the other 'Stimulus'. In fact, in attribution theory it is perfectly possible for either or both participant entities to be persons. Still there is a distinction between the two entities that justifies speaking of generalization across one as involving Consensus whereas generalization across the other is said to involve Distinctiveness. The one and only invariable distinction between the entities is that the Consensus case concerns the entity taking the action or experiencing the state in question. The state or action belongs to the Consensus entity. If a person, then it is an action of the muscles or a state of the mind of the Consensus entity that is to be explained or attributed. The other entity is possibly relevant to that state or action but external to it.

Our sentences all involve two persons and so the best way to map them in attribution terms if the verb names a state is:

$Person_1$	State	$Person_2$
Ted	likes	Paul

If the verb names an action, the mapping is:

$Person_1$	Action	$Person_2$
Ted	helps	Paul

Now the two basic attribution patterns look like this:

Attribution to:	$Person_1$	Attribution to:	$Person_2$
Consensus:	Low	Consensus:	High
Distinctiveness:	Low	Distinctiveness:	High

Let us test the two attribution patterns against intuition. If Ted likes Paul, and few other people like Paul (Low Consensus), and Ted likes many people other than Paul (Low Distinctiveness), then we should attribute the liking of Paul primarily to Ted, which is a $Person_1$ attribution. Specifically, Ted would be the kind of person who likes people; he is amiable or good-natured or whatever. If Ted helps Paul, and few other people help Paul (Low Consensus), and Ted helps many people other than Paul (Low Distinctiveness), then we should attribute the helping of Paul primarily to Ted, which is a $Person_1$ attribution. Specifically, Ted is the kind of person who helps people; he is helpful. By contrast, if Ted likes Paul, and many other people also like Paul (High Consensus), and Ted likes few people other than Paul (High Distinctiveness), then we should attribute the liking of Paul primarily to Paul, which is a $Person_2$ attribution. Specifically, Paul would be the kind of person people like; he is likable. If Ted helps Paul, and many other people help Paul (High Consensus), and Ted helps few other people (High Distinctiveness), then we should attribute the helping of Paul to Paul, which is a $Person_2$ attribution.

Specifically, Paul is the kind of person people help; he deserves or needs or inspires help. All this seems to our intuition to be sound.

In the examples with *like* and *help* you probably noticed that the two schemas and also the actual English derivational adjectives favor one sort of attribution over the other. Since *like* is a state verb, the Stimulus–Experiencer Schema calls for a Stimulus attribution and the Stimulus in the attributional mapping is $Person_2$, Paul, the one external to the state, the one who is likable. That sort of attribution (attribution to $Person_2$) calls for the HH information pattern. *Help* is an action verb and the Agent–Patient Schema calls for an Agent attribution and the Agent in the attributional mapping is $Person_1$, Ted, the one taking the action, the one that is helpful. What would happen if subjects were given our 36 interpersonal interaction sentences and asked not any kind of causal question, but rather asked to say whether Consensus should be High or Low and whether Distinctiveness should be High or Low. If the particular verbs carry, as part of their semantic, distinctive causal weightings, as our work thus far strongly suggests they do, and if subjects have tacit knowledge of the causal inference rules, as many experiments in social psychology strongly indicate, then the verbs should yield predictable combinations of Consensus and Distinctiveness.

We asked the Consensus and Distinctiveness questions indirectly by putting the following task:

> *Ted likes Paul.*
> A.Probably (many – few) other people like Paul.
> B.Probably Ted likes (many – few) other people.

For this problem the Stimulus–Experiencer Schema (as well as the derived form *likable*) predict the answers:

> A.*Many* Generalization across Experiencers (High Consensus)
> B.*Few* Little generalization across Stimuli (High Distinctiveness)

This is a HH pattern which calls for a $Person_2$ (or Stimulus) attribution. To respond 'HH' is really a very indirect way of saying that "Paul is the kind of person people like".

With *Ted helps Paul* the answers to alternatives 'A' and 'B' should be:

> A.*Few* Little generalization across Actors (Low Consensus)
> B.*Many* Generalization across Stimuli (Low Distinctiveness)

This is a LL pattern, which calls for a $Person_1$ attribution. It is an indirect way of saying that "Ted is the kind of person that helps people".

The action verb *help* belongs to Class I (Agent–Patient with Agent as subject), and the prediction for Class I verbs is LL (Agent attributions). The

state verb *like* belongs to Class III (Stimulus–Experiencer with Experiencer as subject), and the prediction for Class III verbs is HH (Stimulus attributions). What is the prediction for Class II verbs such as *charm* and *attract*? The attributional mapping is:

$Person_2$	State	$Person_1$
Ted	charms	Paul

It takes a moment to see that this mapping, initially surprising, is nevertheless correct. $Person_1$ and $Person_2$ are not defined in grammatical terms as subject and object of a sentence let alone in terms of simple linear sequence in a sentence. $Person_1$ and $Person_2$ are not defined in terms of sentences at all but rather in terms of roles played in the interaction depicted by a sentence. $Person_1$, the party to the interaction in relation to whom Consensus is calculated, is definable only as that person experiencing the state to be explained or performing the action whereas $Person_2$ is some relevant external person (a.k.a. the Stimulus). For Class I verbs (e.g., *help*) and Class III verbs (e.g., *like*), $Person_1$ is the grammatical subject of the English sentence depicting the action or event and $Person_2$ the grammatical object. Class II verbs are different. In *Ted charms Paul* or *Ted astonishes Paul* the person experiencing the state, $Person_1$ is the grammatical object and the external person, $Person_2$ (a.k.a. Stimulus), is the grammatical subject. This is the reason why the mapping given above is correct.

Both Class II and Class III verbs are states and so both entail the Stimulus–Experiencer Schema. Therefore, the same prediction must be made in both cases: primary causality assigned the Stimulus and a HH pattern. Notice, however, that because the subject of a sentence depicting a Class II event (e.g., *Ted charms Paul*) is $Person_2$ whereas the object is $Person_1$, the following switch occurs in the meaning of Alternatives A and B:

> Alternative A. Probably (many – few) other people charm Paul. (Distinctiveness)
> Alternative B. Probably Ted charms (many – few) other people. (Consensus)

And so the answers are predicted to be:

> A. *Few* Little generalization across Stimuli (High Distinctiveness)
> B. *Many* Generalization across Experiences (High Consensus)

The rather confusing result of these, regrettably inescapable, reversals of nomenclature is that the pattern HH predicted for both Class II and Class III verbs is, in the case of II, manifest as: A. *Few*; B. *Many*, but in the case of III as: A. *Many*; B. *Few*. And, of necessity, the words A. *Few* and B.

Many as responses to Class II words mean an HH pattern whereas the same words as responses to Class I words mean a pattern of LL. Table 6 summarizes the predictions and presents the results for 20 Harvard–Radcliffe students.

There is one final procedural point. The questionnaire employed used not only the words *many* and *few* but also the equivalent of a 9-point scale. This was done by separating *many* from *few* with just nine dots and treating results as a scale running from *few* (1) to *many* (9). The instructions read:

> This booklet contains a number of simple sentences depicting interpersonal events. For example: *Ted likes Paul.* In connection with each such event, two questions are asked concerning the generality of the event. For instance:
> A. Probably (many . . . . . few) other people like Paul.
> B. Probably Ted likes (many . . . . . few) other people.
> You are asked to estimate the generality of the two aspects of the event by circling one of the nine points between *many* and *few*.

The Alternative A-Alternative B predictions are: Class I, *few–many*; Class II, *few–many*; Class III, *many–few* (Table 6). Subjects were not asked to choose between the words *few* and *many* but rather to encircle one of the nine dots running from *few* (1) to *many* (9) and so the results they produced may be summarized as means. The most straightforward way to translate the predictions expressed in the words *few* and *many* into numerical means is to say that for each class of verbs the *few* prediction should correspond with a lower mean score (therefore nearer the word *few*) than the *many* prediction (nearer the word *many*).

Of the 36 differences between means in Table 6, all but two are in the predicted direction; the means for *enjoy* and *like* are not as predicted but are far from significantly contrary to prediction and in fact they are nearly equal. Twenty-one of 36 differences are significant with $p \leqslant 0.05$ and 13 with $p \leqslant 0.01$. These results for a task that involves mediation by the causal calculus are approximately as strong as those for the simple task of free causal listing. The results must be very striking for anyone who has labored through the immediately preceding explicit description of the causal calculus. The authors find themselves, as practiced as they are, scarcely able to answer their own questionnaire using the explicit calculus. If, however, one is innocent of explicit knowledge (as Ss were) and simply registers the meanings of the sentences, the internalized tacit calculus operates rapidly and rather smoothly to generate the correct answers.

Where do the schemas come from?

This is too deep a question for us, but we will answer it anyway. The results with the causal calculus (Table 6) are suggestive. We know that the various verb patterns imply, and are themselves implied by, one or the other of two informative patterns: HH or LL. These information patterns refer really to the presumed distribution or scope of each predicate 'x' across the human population: High Consensus says that *many* (or *most*) people do or feel 'x' and Low Consensus that *few* people do or feel 'x'. High Distinctiveness says that *few* people are susceptible of being 'x'-ed or capable of inspiring 'x', whereas Low Distinctiveness says that *many* people are susceptible of being 'x'-ed or capable of inspiring 'x'. Our causal calculus questionnaire asked subjects to infer information patterns from verbs, but the implications linking verbs and information patterns are symmetrical and what we want now to suggest is that it is the information patterns or, more exactly, the real differential distribution patterns of the predicates in human beings that give rise to the causal weightings of the various verbs.

Consider the two verbs *flatter* and *slander*. Both are Class I Agent–Patient verbs and both imply the LL pattern, Low Consensus and Low Distinctiveness. Surely this pattern describes the real distribution of flattery and slander across human beings. Just about everyone (most or all persons) can be flattered or slandered. There is no special prerequisite. It is always possible to be the object of slander or flattery, and that is what Low Distinctiveness means. By sharp contrast, however, not everyone, by any means, not even *most* or, perhaps, *many* are disposed to flatter or to slander. This is what Low Consensus means. For verbs like *flatter* and *slander* (Class I verbs generally) it seems to be the case with respect to the predicate 'x' that *many* or *most* people are susceptible of being 'x'-ed whereas relatively fewer people are disposed to do 'x'. In Figure 1(a) this pattern is pictured by showing the Agents (p_1) to be a smaller class than the Patients (p_2).

It follows from the fact that the class p_1 is smaller than the class p_2 that for the purpose of predicting the occurrence or not of a predicate 'x' (e.g., *flattery*) between two unknown individuals, there would be more information (in the mathematical sense) in the knowledge that one participant belonged to the class p_1 (the smaller class, not all humans) than that one participant belonged to the class p_2 (the larger class, almost all humans). To know that one party to an interaction is disposed to flatter is to have some basis for predicting flattery whereas to know only that one party can be flattered is to know little more than that that party is human. We propose that the sense of unequal causal distribution between p_1 and p_2 in Class I verbs derives from the kind of asymmetrical information pictured in Figure 1(a). We propose,

Table 6. *Causal weightings made in the form of consensus and distinctiveness ratings*

Class I	[Agent–Patient [Subject–Object [$Person_1$–$Person_2$	Alternative A Consensus Prediction: Low (*few*)	Alternative B Distinctiveness Prediction: Low (*many*)
apologize to		4.7	5.45
cheat		4.8	7.05**
compete with		5.35	7.8 ***
criticize		3.65	6.75***
defy		4.75	6.5 *
disobey		5.5	5.8
dominate		5.7	6.2
flatter		4.7	6.4 *
harm		4.45	6.2 *
help		3.85	7.4 ***
protect		4.45	5.15
slander		3.55	5.85**
Class II	**[Stimulus–Experiencer [Subject–Object [$Person_2$–$Person_1$**	**Alternative A Distinctiveness Prediction: High (*few*)**	**Alternative B Consensus Prediction: High (*many*)**
astonish		5.4	6.25
attract		4.65	5.6
charm		4.0	6.75***
deceive		3.85	6.7 **
delight		4.4	6.85**
exasperate		5.15	6.8 *
impress		5.05	6.65*
influence		5.55	6.8
repel		3.95	5.8 *
scorn		5.55	5.85
shock		5.65	6.2
trouble		4.8	6.6 *
Class III	**[Experiencer–Stimulus [Subject–Object [$Person_1$–$Person_2$**	**Alternative A Consensus Prediction: High (*many*)**	**Alternative B Distinctiveness Prediction: High (*few*)**
abhor		6.45	4.0 **
admire		5.95	4.5
despise		6.15	4.7 *
detest		5.85	4.8
dread		5.55	3.9
enjoy		6.65	6.8
esteem		6.1	4.0 **
honor		5.7	4.3
like		6.7	6.95
loathe		6.15	3.85**
notice		6.85	4.75**
pity		6.6	4.25**

A > B or B > A: $p \leq 0.05$*, $p \leq 0.01$**, $p \leq 0.001$***.

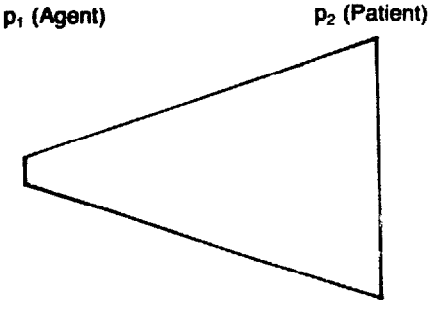
furthermore, that English has created a derived form to name the disposition shared by persons in class p_1 (*flattering*, *slanderous*) precisely because there is enough information in the fact to make the form worth having whereas there would not be for such possible p_2 forms as *flatterable*, *slanderable*, *helpable*, *harmable*, etc.

Verbs of Class II and verbs of Class III are both state verbs to which the Stimulus–Experiencer Schema applies. The only difference between the verbs of II and III is syntactic, subject and object assignment, and this may be ignored now that our goal is to account for the existence of the schema. Consider such verbs as *attract*, *delight*, *astonish*, and *notice*. What percentage of human beings are capable of the experiences in question, capable of being attracted, feeling delight or astonishment, of taking notice? Just about 100% we would think. These are the p_1 persons, the Experiencers, of Figure 1(b). What percentage of human beings are capable of inspiring attraction, delight, astonishment, or notice? Probably also quite a large percentage we would guess, but, considered across both occasions and individuals, surely the ability to inspire affect or notice is less general than the readiness to experience them. It seems so to us at any rate and so we are guessing that the p_2 persons, the Stimulus persons, are fewer than the p_1 persons, as pictured in Figure 1(b). For Stimulus–Experiencer states we propose that the distributional facts, the relative informativeness of classes p_1 and p_2 reverse the pattern of Agent–Patient actions. And so that it is more useful to know that someone is able to attract, delight, astonish, or be noticed than it is to know that someone is able to feel attraction, delight, or astonishment, or to take notice. And, furthermore, that the dispositions attractive, delightful, astonishing, and noticeable are worth naming whereas the dispositions attractable, delightable, astonishable, and noticing are not.

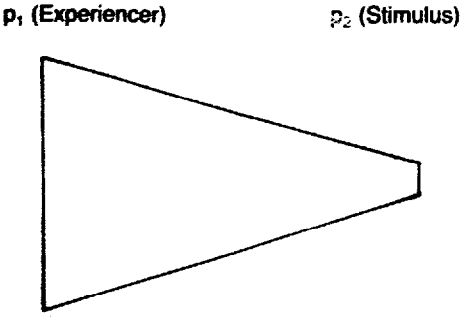
If actions and states do differ as suggested in Figure 1, the very hard question remains: "Why do they so differ?" Our answer to this one must be admitted to be tentative. The claim is a simple one. For both types of verb (Stimulus–Experiencer and Agent–Patient) we will give reasons why one verb argument, the argument applying to the larger class of persons, should tend to be true of humans generally whereas for the other verb argument, the argument applying to the smaller class of persons, we will give reasons why it ought only to apply to subclasses of humans. The less general arguments are probably so because the dispositions in question depend on individual, non-uniform learning histories and upon genetic variation. The more general argument in the Stimulus–Experiencer case is on the Experiencer side, and it is probably general because it is innate. The more general argument in the Agent–Patient case is not a response (or experience) and cannot be general because it is innate. It seems to us that these Patient dispositions must tend

Figure 1. *Relative sizes of classes of persons to whom predicate arguments apply.*

a) Class I Agent–Patient Verbs



b) Class II and Class III Stimulus–Experiencer Verbs



to be very general just because they belong to the human condition. Everyone can be flattered, slandered, helped, harmed, cheated, and disobeyed because that is the way it is with humans in society; these are not innate propensities but passive, unavoidable susceptibilities.

Let us think first of the experiences or states, the meanings of the verbs in Class II and Class III. There are 24 verbs in all, but most of them name some shade of either positive affect (e.g., *charm*, *attract*, *delight*, *admire*) or of negative affect (e.g., *trouble*, *exasperate*, *detest*, *dread*). While the various shades of negative and positive affect (compounded with cognition) that the individual verbs name can hardly be innate in humans, the core negative and positive affects themselves can be considered innate. This is to say that these feelings are 'wired in' to the extent that all humans, in the range of environments in which humans now develop, have positive and negative feelings and have them from infancy. If the feeling states named by the verbs of Class II and III are largely innate and so can be reliably found in all human beings,

then the class p_1 for those states is a very large class, just as it is pictured to be in Figure 1(b).

Some verbs in Class II and Class III do not name feelings. *Astonish* and *notice* (perhaps *shock*) name involuntary cognitive reactions of arousal and attention. These reactions (*surprise-surprising* would be another) are like positive and negative affect in that they are, as reactions, innate in humans and universal in the species. To know that someone is capable of feeling surprise or of taking notice is to know no more than that he is human, and so the representation of class p_1 in Figure 1(b) seems as appropriate for the arousal verbs as for the affect verbs.

If all human beings are roughly equivalent in their capacity to feel positive and negative affect and to experience arousal and attention, it is not the case that all human beings are equivalently adequate as stimuli for eliciting these states. The equivalence on the response side seems not to be matched on the stimulus side. There are individuals who attract and individuals who repel; there are individuals who are noticeable and individuals who are not. This kind of variation would seem to be partly genetic in origin (physical features and temperament) and partly attributable to cultural learning (conceptions of beauty, decorum, etc.) and partly due to idiosyncratic learning histories. In any case, there is nothing like a constant level of charm, attractiveness, delightfulness, or noticeability uniform in the species. Therefore, it is informative to know of a given individual which predicate arguments apply and which do not, and to communicate that information English must have and does have the relevant words. The classes of human beings to whom the various predicate arguments apply (p_2) should be relatively small, and that is the way they are pictured in Figure 1(b).

For action verbs (Agent–Patient) the responses named (e.g., *cheat, compete with, flatter, slander, help*) are certainly not innate but are the kinds of responses we think of as typical products of genetic variation, socialization, and idiosyncratic learning histories. Individual variation is taken for granted and, indeed, many of the personality variables of scientific psychology are agentive dispositions. The other argument of the action verb, the Patient, is an essentially passive role: to be cheatable, flatterable, helpable, etc. It seems to us that individual variation is not great for Patient dispositions simply because it is the human lot to be able to be acted upon in these ways. If one is human, one can be harmed—also helped. The suggestion is, then, that the class of p_2 persons (Patients) in Figure 1(b) is large because the arguments in question are the passive susceptibilities of our species.

Summary

We have discovered several things about the English verbs that name interactions, either mental or behavioral, between two persons; verbs such as *like*, *charm*, *notice* (mental) and *help*, *cheat*, *criticize* (behavioral) in such a context as *Ted* —— *Paul*. The most general discovery is that adult native speakers of English think of causality in such inter-personal interactions as unequally apportioned between the interactants. Four different methods are used to inquire about the apportioning of causality between interactants (e.g., *Ted* and *Paul*): 1) Direct causal ratings with simple sentences in the active voice; 2) Direct causal ratings with simple sentences in the passive voice; 3) Free listing of causes of the interaction; 4) Indirect ratings mediated by the calculus of causality. The four methods yield the same highly consistent results.

For action verbs that name voluntary behavior (e.g., *help*, *cheat*, and *criticize*) greater causal weight is given to the Agent argument of the verb (e.g., *Ted* in *Ted helps Paul*) than to the Patient argument (e.g., *Paul*). For state verbs that name mental states or experiences (e.g., *like*, *charm*, *notice*) greater causal weight is given to the Stimulus argument of the verb (e.g., *Paul* in Ted likes Paul) than to the Experiencer argument (e.g., *Ted*). In addition to the unequal distribution of causal weights characteristic of the two classes of verbs (Agent–Patient; Stimulus–Experiencer), the individual verbs, within each class, tend to manifest a specific balance of causality that is characteristic of the verb across subjects and across the four methods of inquiry.

For English verbs of the type we have studied, derivational adjectives also often exist (e.g., *likable*, *charming*, *noticeable*, *helpful*, *cheating*, *critical*). For the 36 English verbs we have studied (using four methods of inquiry), it is invariably the case that the verb argument assigned the greater causal weight (Agent or Stimulus) is the argument to which the corresponding derivational adjective is attributive. In *Ted helps Paul* the Agent, *Ted*, is assigned the greater causal weight and the derived adjective *helpful* is attributive to *Ted* or may be said to name a disposition in *Ted*. In *Ted likes Paul* the Stimulus, *Paul*, is assigned the greater causal weight and the derived adjective *likable* is attributive to *Paul* or may be said to name a potential in *Paul*. For the 36 verbs studied there are in English no existent derived adjectives (defined by a frequency criterion, not by grammatical possibility) attributive to Patient or Experiencer. This relation between derivational morphology and causal weighting, which is true without exception of the 36-verb sample studied, has been shown to be very generally, though not exceptionlessly, true of the total population of interaction verbs in English. In fact, it looks as if English derivational morphology in this domain follows the rule: Create adjectives from action verb bases that are attributive to the Agent and adjectives from

state verb bases that are attributive to the Stimulus.

Two kinds of processing model come to mind for the causal weighting performance (in all four versions). A priming model would work by activating the particular derivational adjective linked to its verb base in a semantic network. A causal schema model would not involve the particular adjective at all but would, instead, operate with lexical entries for verbs which included causal weighting information. This information might simply specify the argument assigned primary causal weight (i.e., Agent or Stimulus), but the evidence that individual verbs within a class tend to have characteristic specific causal balances suggests that the lexical entry might specify a quantitative apportioning of causality as between Agent and Patient, Stimulus and Experiencer. Research results may eventually choose between a priming model and a causal schema model, but as yet there is no reason to favor one over the other.

Because certain facts about English morphology predict certain ways of thinking about causality in verbs (operationalized in four ways), our main finding may seem to be a Whorfian one, a demonstration that language affects thought. We think it is not that but is, rather, a demonstration that a mode of thought that is universally human affects language use. We have at present only one good reason for favoring the latter formulation. A distinction between derivational adjectives that do and do not exist in English cannot, in fact, be made in *linguistic* terms, that is in terms of a lexicon or grammar. The generative morphological system of English is not biased in favor of adjectives attributive to Agent and Stimulus and against adjectives attributive to Patient and Experiencer. All types are equally possible—grammatically. The distinction can be drawn in terms of frequency. Of the four varieties of possible derivational adjectives, it is those attributive to Agent and Stimulus (for the most part) that satisfy a minimum frequency criterion. Frequencies are usage, and usage is the way people deploy a language, not what a language system permits. Therefore, we think it likely that our main finding is to be explained as follows. Humans generally think in terms of the Agent–Patient Schema and the Stimulus–Experiencer Schema. Those humans whose first language is English have used the resources of English in a way that reflects the two schemas and so there seem to be words for Agent dispositions but not for Patient dispositions, words for Stimulus dispositions but not for Experiencer dispositions. And subjects assign primary causal weights to the predicate arguments that the adjectives modify.

Assuming that it is correct to suppose that the two schemas are universals of human thought, we have speculated about why they should be so. It seems to us that the Patient argument of action verbs applies to almost everyone because it is a part of the passive human condition whereas the Agent argu-

ments seem only to apply to subclasses of humans because they are dispositions built up by individual genetics and learning histories. The Experiencer arguments seem to us to apply to all humans because they are innate responses whereas the Stimulus arguments apply only to subclasses having distinctive genetic and acquired characteristics. Such speculation is perhaps premature, however, because there is not as yet any strong evidence that the causal schemas are universals of human thought. In search of that evidence or of contrary evidence, we turn next to languages unrelated to English.

References

Allport, G.W., and Odbert, H.S. (1936) Trait names: A psycho-lexical study. *Psychol. Monogr.*, XLVII, Whole No. 211.

Aronoff, M. (1976) Word formation. *Linguistic Inquiry Monograph One*. MIT Press.

Brown, R. (1958) *Words and Things*. Glencoe, Illinois, Free Press.

Brown, R., and Lenneberg, E.H. (1954) A study in language and cognition. *J. Abnorm. Soc. Psychol.*, *49*, 451–462.

Caramazza, A., Grober, E., Garvey, C., and Yates, J. (1977) Comprehension of anaphoric pronouns. *J. Verb. Learn. Verb. Behav.*, *16*, 601–609.

Chafe, W.L. (1970) *Meaning and the Structure of Language*. Chicago, The University of Chicago Press.

Comrie, B. (1981) *Language Universals and Linguistic Typology*. Chicago, The University of Chicago Press.

Cunningham, J.D., Starr, P.A., and Kanouse, D.E. (1979) Self as actor, active observer, and passive observer: Implications for causal attributions. *J. Person. Soc. Psychol.*, *37*, 1146–1152.

Fillmore, C.J. (1968) The case for case. In E. Bach and R.G. Harms (eds.), *Universals in Linguistic Theory*. New York, Holt, Rinehart, and Winston, 1–87.

Francis, W.N. (1958) *The Structure of American English*. New York, Ronald Press.

Garvey, C., and Caramazza, A. (1974) Implicit causality in verbs. *Linguistic Inquiry*, *5*, 459–464.

Garvey, C., Caramazza, A., and Yates, J. (1976) Factors influencing assignment of pronoun antecedents. *Cog.*, *3*, 227–243.

Heider, F. (1958) *The Psychology of Interpersonal Relations*. New York, Wiley.

Jones, E.E., and Nisbett, R.E. (1972) The actor and the observer: Divergent perceptions of the causes of behavior. In E.E. Jones, D.E. Kanouse, H.H. Kelley, R.E. Nisbett, S. Valins, and B. Weiner (eds.), *Attribution; Perceiving the Causes of Behavior*. Morristown, N.J., General Learning.

Kelley, H.H. (1967) Attribution theory in social psychology. *Nebraska Symposium on Motivation*, *15*, 192–238.

Kelley, H.H., and Michela, J.L. (1979) Attribution theory and research. *Annual Review of Psychology*, *31*, 457–503.

Kučera, H., and Francis, W.N. (1967) *A Computational Analysis of Present-Day American English*. Providence, Brown University Press.

McArthur, L.Z. (1972) The how and what of why; Some determinants and consequences of causal attributions. *J. Person. Soc. Psychol.*, *22*, 171–188.

Miller, G.A., and Johnson-Laird, P.N. (1976) *Language and Perception*. Cambridge, Mass., Harvard University Press.

Orvis, B.R., Cunningham, J.D., and Kelley, H.H. (1975) A closer examination of causal inference: The roles of consensus, distinctiveness, and consistency information. *J. Person. Soc. Psychol*, *32*, 605–616.

Oxford English Dictionary on Historical Principles (1933) Edited by J.A.H. Murray, H. Bradley, W.A. Craigie, and C.T. Onions. Oxford, Clarendon Press.

Random House Dictionary of the English Language; The Unabridged Edition, 1966.
Ross, L. (1977) The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (ed.), *Advances in Experimental Social Psychology*. Volume 10. New York, Academic Press.
Skinner, B.F. (1938) *The Behavior of Organisms*. New York, Appleton-Century-Crofts.

Résumé

L'étude porte sur la causalité implicite des verbes anglais qui désignent des interactions mentales ou comportementales entre deux personnes. Ce sont des verbes du type: aimer, remarquer (mental) ou aider (comportemental) dans un contexte tel que Ted Paul. En utilisant quatre méthodes différentes, on montre que des locuteurs adultes considèrent que la causalité dans de tels verbes est inégalement distribuée entre les interactants. Pour les verbes de comportement ou d'action un poids causal plus important est donné à l'argument Agent du verbe (c'est à dire *Ted* dans *Ted* aide *Paul*) qu'à l'argument Patient (*Paul*). Pour les verbes mentaux ou d'état un poids causal plus important est donné à l'argument Stimulus du verbe (Ex., *Paul* dans *Ted* aime *Paul*) qu'à l'argument Expérimentant (*Ted*). Pour les verbes anglais du type étudié, il existe souvent des adjectifs dérivationnels (Ex., aidant, aimable, remarquable). Ces adjectifs s'attribuent à l'un ou l'autre des arguments des verbes (Agent ou Patient, Stimulus ou Experimentant). On montre que la direction de l'attribution causale dans l'adjectif (Ex., aidant est attributif à *Ted* l'Agent, aimable est attributif à *Paul* le Stimulus) prédit le poids causal primaire assigné dans nos tâches expérimentales. On montre également que les adjectifs anglais dérivés des verbes d'action sont toujours attributifs à l'Agent alors que les adjectifs dérivés des verbes d'état le sont pour le Stimulus. Comme certains faits sur la morphologie de l'anglais prédisent certaines façons de penser la causalité, l'apport principal de l'étude semble être Whorfienne et démontrer que le langage affecte la pensée. Cependant, il est argumenté que tel n'est pas le cas et démontré que deux modes de pensée (le schéma Agent–Patient et le schema Stimulus–Expérimentant) affectent l'usage de la langue. Il est suggeré que les schemas sont des universaux de la pensée humaine.