Thematic Roles as Verb-specific Concepts

Ken McRae, Todd R. Ferretti and Liane Amyote

University of Western Ontario, London, Ontario, Canada

Thematic roles are typically viewed as slot and filler mechanisms in which the slots are devoid of content, or the content is limited to a few syntactically relevant features. This traditional view excludes important knowledge that people possess about who tends to do what to whom in specific situations. Thus it was extended here by treating thematic roles as verb-specific, feature-based concepts. In Experiment 1, subjects produced features for agent and patient roles such as “someone who is frightened”. Role/filler typicality ratings, in which subjects provided judgements for questions such as “How common is it for a monster to frighten someone?” (i.e. monster as an agent of frighten), were collected in Experiment 2A. In Experiment 2B, different subjects were presented with the role features from Experiment 1 and the typical agents and patients from Experiment 2A. They were asked to rate how central each of the role features was to the nominal concepts, and role/filler featural similarity was calculated from these ratings. Analogous to studies of object concepts, role/filler featural similarity predicted role/filler typicality, thus suggesting that thematic roles and nominal concepts are represented in similar forms. Experiment 3 was a self-paced reading study in which adjectival features were used to bias an initial noun concept towards being a good agent or patient of the past participle of a reduced relative clause (“The shrewd heartless/young naive gambler manipulated by the dealer had bid more than he could afford to lose”). Featural bias modulated ambiguity resolution. The present work, in conjunction with other recent results, suggests that thematic roles might best be viewed as verb-specific concepts, and that this conceptual/world knowledge is computed and used immediately in on-line language processing. Results are discussed in relation to other recent treatments of thematic roles as well as constraint-based and garden path models.

Requests for reprints should be addressed to Ken McRae, Department of Psychology, Social Science Centre, University of Western Ontario, London, Ontario N6A 5C2, Canada. E-mail: kmm@sunrae.ssc.l.uwo.ca

This work was supported by NSERC Grant OGP0155704 to the first author. A subset of this research formed part of T.R.F.’s Master’s thesis. Parts of this research were presented at the 1994 Meeting of the Psychonomic Society and the 1995 CUNY Sentence Processing Conference. We are indebted to Michael Spivey-Knowlton, Mike Tanenhaus, Amit Almor and two anonymous reviewers for illuminating discussions and/or helpful comments on earlier drafts. The complete set of feature norms from Experiments 2 and 3 can be obtained from K.M.

© 1997 Psychology Press, an imprint of Erlbaum (UK) Taylor & Francis Ltd
INTRODUCTION

It is well known that people typically understand language with little if any delay (Marslen-Wilson, 1975). Consistent with this observation, recent research has demonstrated that on-line language comprehension involves quickly combining a number of informational constraints to arrive at an interpretation that best satisfies them. Some of the constraints that have been investigated to date are syntactic principles (Frazier & Rayner, 1982), lexically specific syntactic knowledge, including both the relative frequency of verb forms (MacDonald, Pearlmutter, & Seidenberg, 1994) and subcategorisation frames (Trueswell, Tanenhaus, & Kello, 1993), accrued information about the discourse (Altmann, Garnham, & Dennis, 1992; Spivey-Knowlton & Sedivy, 1995) and relevant aspects of the visual environment (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). The aim of this paper is to further this line of research by focusing on an important lexical-conceptual constraint, namely thematic roles (sometimes called case roles, thematic relations or semantic case, although these terms can refer to slightly different notions). There exists a considerable body of evidence demonstrating that people’s thematic role knowledge is used quickly to constrain sentence interpretation. Evidence has been found in studies of the main clause/reduced relative clause ambiguity (McRae, Spivey-Knowlton, & Tanenhaus, in prep.; Pearlmutter & MacDonald, 1992; Trueswell, Tanenhaus, & Garnsey, 1994), the prepositional phrase ambiguity (Taraban & McClelland, 1988), the clause boundary ambiguity (Clifton, 1993; Stowe, 1989) and filler–gap constructions (Boland, Tanenhaus, Garnsey, & Carlson, 1995). Consequently, thematic role knowledge is a central component of most current theories of on-line sentence comprehension (Carlson & Tanenhaus, 1988; Frazier, 1987; MacDonald et al., 1994) and language comprehension in general (Graesser, Singer, & Trabasso, 1994). The present work outlines and tests a conceptually based theory of thematic roles that has direct implications for on-line sentence interpretation and the relationship between linguistic and world knowledge.

Thematic roles have been defined as the “semantic roles that may be played by the subcategorized complements (or arguments) of a verb” (Tanenhaus, Carlson, & Trueswell, 1989, p. 212). Roughly, they represent the roles that participants (widely defined) play in the events described by verbs. Linguists have proposed numerous thematic taxonomies that typically include roles such as agent, patient, theme, goal, instrument and location (e.g. Cook, 1979; Dowty, 1991; Fillmore, 1968). Three of these are illustrated in the sentence, “Mary ate the pizza with her hands”. In this sentence, “Mary” is the agent because she is performing the event, “the pizza” is the patient because the event is being performed on it, and “her
hands” are the instruments because they are being used to perform the event.

The research reported below examined thematic roles by focusing on agents and patients. The precise definition of the agent and patient roles differs widely across researchers. Our choice of nomenclature was influenced by Jackendoff (1987, p. 378), who stated that “thematic relations are to be reduced to structural configurations in conceptual structure; the names for them are just convenient mnemonics for particularly prominent configurations” (see Ladusaw & Dowty, 1988, for a similar view). In other words, because the particular labels given to roles are not crucial in our framework, “agent” and “patient” are used simply as convenient mnemonics for the entity performing the event and the entity having the event performed on it, respectively. Note that this differs from theories in which roles or cases are hypothesised to be conceptual primitives, so that special significance is attached to the labels, as in Fillmore (1968).

THE CONTENT OF A THEMATIC ROLE

Thematic roles have been studied intensively in linguistics and psycholinguistics for about 30 years. During this time, a number of viewpoints have been expressed concerning whether they are syntactic, semantic or conceptual in nature (e.g. Dowty, 1991; Fillmore, 1968; Jackendoff, 1972; Schlesinger, 1995). Some theorists claim that the content of a thematic role is limited to a few syntactically relevant selectional restrictions (e.g. animacy; Caplan, Hildebrandt, & Waters, 1994; Chomsky, 1965), whereas others have defined a role as a set of a small number of general features (Dowty, 1991) or conceptually primitive dimensions such as cause and control (Schlesinger, 1995).1 The main issue addressed in this paper concerns the amount and type of content allotted to thematic roles. This issue is important because theoretical differences in thematic role content have direct implications for the on-line assignment of noun phrases (NPs) to possible roles, a process that is viewed as central to language comprehension. In addition, this issue has direct implications for the relationship between linguistic and world knowledge. For example, theories from the Chomskyan perspective draw sharp boundaries between syntax and semantics, so that thematic roles are viewed as syntactically specified slot and filler mechanisms with minimal conceptual content (e.g. Caplan et al., 1994; Schlesinger, 1995). Consequently, on-line thematic assignment is

---

1 In this paper, a feature is enclosed in angled brackets, a concept name is printed in upper case when it is meant to denote the mental representation of the concept, experimental examples are in upper case also, and experimental sentences or sentence fragments are enclosed by quotes when they occur in the text (in contrast to being set off by themselves and numbered).
primarily driven by syntactic factors such as word order or the phrase’s position in a phrase-structure tree or subcategorisation frame. World knowledge, such as the fact that judges typically convict other people but are rarely convicted themselves, is viewed as stored separately from linguistic knowledge and is used for checking the accuracy of thematic assignments, but not for guiding them. In contrast, in this paper, we explore the notion that thematic roles are primarily conceptual, reflecting world knowledge about specific events and the entities and objects that participate in such events. Furthermore, we claim that this type of world knowledge is organised in a manner that allows it to be computed and used immediately in language processing. Predictions derived from these notions are tested in three experiments.

The present view borrows from a number of treatments of thematic roles. For example, Tanenhaus et al. (1989) argued that thematic roles are lexical-conceptual entities with close ties to syntactic information. They also claimed that the act of recognising a verb makes available information about the sense of the verb (i.e. its core meaning), its thematic roles, the types of constituents that can be complements, and information concerning how the roles and constituents are interconnected. Thus, as in Jackendoff (1987), thematic role knowledge is viewed as lexical, in that it is part of a verb’s meaning, and this knowledge is assumed to be automatically made available when a verb is read or heard. Along these lines, Gentner (1981) has argued that verbs are relational concepts and the present emphasis on thematic role concepts can be construed as a claim that an important aspect of a verb’s meaning is its relation to the entities and objects that commonly participate in the events described by that verb. In other words, the meaning of many verbs is carried by the specification of the action taking place in conjunction with information about the common participants. In contrast, a number of theories of thematic roles (particularly in the form of case or theta roles) have treated them as part of the grammar—that is, as syntactic knowledge (e.g. Culicover, 1988). This is particularly evident in theories that treat them as slot and filler mechanisms with minimal content (i.e. most treatments from the Chomskyan tradition), a position that often includes the view that roles are innately specified, non-decomposable primitives. For example, Schlesinger (1995, p. 5) espouses a view of this sort, claiming that cases are regarded as fuzzy categories that are “linguistic constructs that function in the grammar”. Schlesinger was careful to distinguish them from cognitive “notions” about what types of people and things might be good fillers for roles. Finally, in the artificial intelligence literature, where thematic roles have played an important part in a number of implemented parsing models, this issue has been handled in a number of ways. Perhaps Cottrell (1988, p. 171) said it best when he asked “what counts as ‘syntactic’ and what ‘semantic’? The answer is usually a matter of taste”.
The present work also takes a lead from Fillmore (1968), who claimed that cases exist in cognition and somewhat independently of language, and that this knowledge is used efficiently during language processing. In addition, we draw no sharp distinctions among lexical, semantic, conceptual and episodic knowledge, as in Jackendoff (1983), but contrary to Schlesinger (1995). Verbs describe episodes; that is, one way to view them is as labels for sets of similar events (i.e. event categories; Kersten & Billman, 1995). Thus a verb’s semantic representation is computed from the set of episodes that have been linked to the verb’s word (i.e. spelling or sound). In what can be take as support for these claims, McKoon, Ratcliff and Dell (1986) have persuasively argued that there is little if any evidence to support Tulving’s (1972, 1983) distinction between semantic and episodic memory, arguing instead that word meaning and general world knowledge is stored in the same form as episodes, and presumably constructed on-line from them. The next section provides a description of one aspect of verb meaning, namely, the notion of a thematic role concept.

Computed Prototypes

In a recent treatment of thematic roles, Dowty (1991) claimed that the search for the optimal taxonomy of roles was at an impasse, and thus concluded that it was time to discard the view that roles are discrete and categorical. To remedy this situation, he borrowed from prototype theories of concepts and categorisation to propose that thematic roles consist of lists of features (Rosch and Mervis, 1975). Dowty posited two basic roles, proto-agent and proto-patient. He described the proto-patient role, for example, as having the following general features: \( \langle \text{undergoes change of state} \rangle; \langle \text{incremental theme} \rangle; \langle \text{causally affected by another subject} \rangle; \langle \text{stationary relative to movement of another subject} \rangle; \text{and} \langle \text{does not exist independently of the event or not at all} \rangle. \) According to Dowty, the patient role for a specific verb constitutes a proper subset of these features. It should be noted that other linguists such as Schlesinger (1995) have stressed that role features of this type are innate and/or conceptually primitive, although Dowty did not. We took Dowty’s lead and moved a step further. Our view resembles Dowty’s in that each role of each verb is a concept. However, rather than being a proper subset of Dowty’s general features, role concepts are formed through the everyday experiences during which people learn about the entities and objects that tend to play certain roles in certain events. This knowledge then serves as an important constraint in on-line language comprehension and production. For example, a person’s representation of the agent role of ACCUSE is a result of her experiences with people who accuse others in everyday occurrences and in linguistic descriptions of them.
Via a computed role concept, this knowledge is made available when ACCUSE is read or heard, and hence is a key factor in driving on-line thematic assignment. (A discussion of the implications for on-line thematic assignment is deferred until the introduction to Experiment 3, an on-line reading study.)

What form might these role concepts take? When dealing with lexical concepts, it is important to distinguish between the information that is laid down in memory and the representation that is formed on-line when word meaning is computed. For example, some researchers have claimed that specific instances or episodes are encoded in memory (e.g. Hintzman, 1986), whereas others have stressed that knowledge is an abstraction over those instances (e.g. Posner & Keele, 1968). In distributed connectionist models such as that of McClelland and Rumelhart (1985), specific episodes are encoded in a superimposed memory system. In a system such as this, whether a computed representation corresponds to a prototype versus a specific episode depends on the amount and quality of relevant context that is included in the memory probe; that is, the better the probe context matches the context of a specific episode, the greater is the likelihood of recalling it. Given a memory system like that of Hintzman or McClelland and Rumelhart, when a word is read or heard, the resultant representation that is computed on-line is akin to a feature-based prototype. For example, when someone reads or hears a word like DOG in the absence of distinguishing context, she computes a representation that consists of the features typically possessed by the dogs that she has encountered and encoded (e.g. has paws, barks, wags tail, etc.). Likewise, a computed role concept can be viewed as a set of features that is typically possessed by the fillers of that role. Take, for example, the agent role of ACCUSE. Most people have been accused of something or have witnessed others being accused. Although a wide range of individuals may be accusers at different times, they often possess certain features, such as being mean, judgemental or insecure, and it is features like these that are most likely computed as part of the agent role representation of the verb.

Note that prototype theory is not being advocated here in its most restricted sense. In the type of system envisioned, the features and their relative saliences that comprise a computed role concept differ over time as a result of experience with events and ones similar to it. Barsalou (1987) has conducted a series of studies to illustrate that the internal structure of categories (as operationalised by typicality ratings) differs between individuals as well as within individuals over time. Furthermore, the computed representation differs with context, so Barslou’s (1982) notions of context-dependent and context-independent features are relevant here. Finally, analogous to the claims of Rosch (1978), Malt and Smith (1984) and McRae, de Sa and Seidenberg (1997), as in the realm of object concepts, it is
believed that a complete understanding of thematic role concepts necessitates an exploration of how features correlate both within and across roles, a factor that is excluded from basic prototype theories (and is not investigated in this paper).

A theory that posits computed prototype-style representations is advantageous for a number of reasons. First, prototype theory is pertinent to language processing in that reading or hearing a word appears to result in the activation of some sort of prototype (see Hintzman, 1986, for a particularly enlightening demonstration of this). This is particularly true in psycholinguistic experiments, in which context is typically absent or impoverished. Second, prototype theory has successfully been used to account for human performance in numerous studies of lexical concepts (e.g. Rosch & Mervis, 1975; Tversky, 1977). Third, prototype theory entails that concepts have fuzzy boundaries and internal structure. Fuzzy boundaries for role concepts are necessary to capture the fact that it is not always clear whether or not a filler is appropriate for a role. Although there are cases where the boundaries are quite clear (e.g. the agent role of EAT must be filled by an animate being or the meaning becomes metaphorical, as in “That photocopier ate my memo”), there are many verbs for which people would disagree about the appropriateness of boundary examples (see McCloskey & Glucksberg, 1978, for corresponding experiments with concrete nouns). Even more central to the experiments presented below is the fact that prototype theory directly implies that concepts have internal structure (e.g. a chair is a better example of a piece of FURNITURE than is a lamp or a rug). When applied to thematic roles, it leads to the prediction that some nouns fill a role better than others. For example, a waiter is a great agent of SERVE, whereas a nurse is a good one. Furthermore, although a reporter is far from a prototypical agent, a reporter can plausibly bring someone their lunch or a cup of coffee. Finally, prototype theory is attractive because a prototype need not correspond to any single lexical concept. Thus this notion automatically entails the view that activating a thematic role concept is not akin to activating a set of words, but instead is a process of activating features that overlap more or less with a number of lexical concepts.

Verb-specific Features

The type of feature employed differentiates our approach from previous ones. Experiment 1 investigated role concepts by using a norming procedure similar to those used in numerous studies of object concepts (e.g. Rosch & Mervis, 1975). Specifically, subjects listed features for agent and patient roles such as “someone who is served”, thus providing detailed information about the roles of specific events. This approach contrasts with those of Schlesinger (1995) and Dowty (1991), who used their intuition to derive sets
of conceptual primitives or general features. Note that no subject listed this type of feature in Experiment 1. The empirically derived features can also be compared and contrasted with selectional restrictions that have been incorporated into a number of theories of thematic and case roles (e.g. Chomsky, 1965; Fillmore, 1968). Theories vary in terms of the number and complexity of selectional restrictions used. Some incorporate a small number of syntactically relevant restrictions such as animacy (Caplan et al., 1994), whereas other theories, particularly artificial intelligence parsers, incorporate a large number of rather elaborate ones (e.g. Cottrell, 1988). In a sense, the present work could be viewed as positing elaborate, empirically derived, event-specific selectional restrictions. There is a key difference, however. Selectional restrictions have a categorical feel to them that can be viewed as roughly corresponding to a classical theory of concepts (see Smith & Medin, 1981, for a description). Thus a selectional restriction specifies a condition that a filler must satisfy to avoid creating an anomalous or metaphorical sentence. This property clashes directly with a central tenet of prototype theory; namely, that features are characteristic rather than defining.

HIGHER-LEVEL GENERALISATIONS

It should be noted that our aim is to elaborate, rather than replace, theories that treat roles as slot and filler mechanisms or those that use general features. A verb-general notion of thematic roles has played an important part in discovering linguistic generalisations and specifying rules that link roles to grammatical positions like subject and direct object (Fillmore, 1968; Jackendoff, 1972), although the utility of linking rules has recently come under attack (Gropen, Pinker, Hollander, & Goldberg, 1991). Furthermore, verb-specific thematic role concepts might relate to higher-level verb-general roles in systematic ways. Most theories of object concepts rely in some way on the notion of a hierarchical structure of categories, as illustrated by the following set of concepts: robin → bird → animal → living thing. An analogous hierarchical structure of role categories is envisioned that is based on semantic neighbourhoods of verbs (e.g. “smash” at the basic level → break-type verbs → verbs of aggression → action verbs). Presumably, the general features discussed by Dowty (1991) and others are important to describing generalisations that are related to the higher levels in a hierarchy of this sort. To illustrate this, consider the effects of various levels when someone attempts to understand a sentence such as “The pencil kicked the cow”. Because English word order strongly constrains interpretation, to make sense of it, the comprehender must contract features from the agent role concept to PENCIL. That is, the comprehender must
imagine that the pencil is \(\langle\text{alive}\rangle\) and \(\langle\text{can move}\rangle\) (verb-general), that it \(\langle\text{has legs}\rangle\) (probably verb-specific), and that maybe it is \(\langle\text{mean}\rangle\) or \(\langle\text{angry}\rangle\) (verb-specific). A further verb-specific aspect is revealed by contrasting KICKED with PUNCHED; understanding the above sentence does not require imagining that the pencil \(\langle\text{has arms}\rangle\), although it would if KICKED was replaced by the related verb PUNCHED.

PREDICTIONS

The experiments presented below tested five main hypotheses derived from our view of thematic roles as feature-based, verb-specific, computed prototypes. In Experiment 1, subjects listed features for specific roles such as “someone who is convicted” and these featural descriptions were used to create first-order approximations to 20 role concepts that served as the basis for Experiment 2. The representations were used to test whether some roles are better defined than others (presumably because there are key differences in the regularities among the entities or objects that form the basis of the role concept). In Experiment 2A, subjects rated role/filler typicality. This involved indicating on a 7-point scale how common it is for some concept, such as CRIMINAL or JUROR, to be an agent or patient in an event described by a verb (e.g. “How common is it for a criminal to convict someone?”). These data were used to test the second hypothesis; namely, whether role concepts have internal structure, as predicted by prototype theory. Experiment 2B tested the hypothesis that this internal structure might be accounted for by role/filler featural similarity, thus testing whether thematic roles are like nominal concepts in at least one important respect. In Experiment 2B, subjects were presented with the features from Experiment 1 and rated their importance to concepts that Experiment 2A had established as either good agents or patients. Using a variation on the logic that has been applied to studies of object concepts (e.g. Malt & Smith, 1984; Rosch & Mervis, 1975), role/filler featural similarity was used to account for role/filler typicality. Finally, Experiment 3 tested two further hypotheses. First, it tested whether the thematic role knowledge studied in Experiments 1 and 2 is linguistically relevant in the sense that it is computed and used immediately and automatically in on-line language processing. In so doing, it tested whether a subtle featural manipulation would have an early effect on syntactic ambiguity resolution, a prediction relevant to the current debate between constraint-based and garden path theories of sentence comprehension. In Experiment 3, features (in the form of adjectives) were used to bias the initial NP of a sentence towards being a good filler for either the agent or patient role of the verb, and this featural information modulated resolution of the main clause/reduced relative clause ambiguity.
EXPERIMENT 1

Subjects’ knowledge of role concepts was tapped by asking them to generate features that were then used to construct conceptual representations. This feature-listing method has been used in many previous studies (e.g. Barsalou, Olseth, & Wu, 1995; Rosch & Mervis, 1975; Smith, Osherson, Rips, & Keane, 1988). The resulting norms are assumed to provide valid information not because they yield a literal record of semantic representations, but rather because such representations are systematically used to generate features. Thus feature norms provide a window into important aspects of concepts without necessarily being definitive (Medin, 1989). The agent and patient roles of 20 verbs were normed. Subjects were given a form containing one role per verb (e.g. “someone who is convicted”) and were asked to list features for them (e.g. \(\text{K is guilty}\), \(\text{K has broken the law}\)).

There were three purposes of Experiment 1. The first was to test whether subjects would be able to provide coherent lists of features for roles such as “someone who is convicted”. The working assumption was that they would. Thus the second purpose was to use the feature lists to construct representations of role concepts in terms of independent feature lists. Although role concepts are undoubtedly more complex than lists of individual features, this representation was used as a starting point for our investigation. The third purpose was to test whether differences existed among roles in terms of the ease with which subjects could produce features for them. Roles differ in terms of the distribution of exemplars that fill them, and these differences may have implications for the corresponding computed representations. Three aspects of the distribution of exemplars in a thematic role category may have important repercussions for the featural representation of that role: the number of exemplars that commonly fill the role; shared features among the common exemplars (inter-exemplar featural similarity); and the number of exemplars that can plausibly fill the role. Consider the agent roles of the verbs CONVICT and ACCUSE. There are only a few common agents of convicting events in our society, and these same agents are also the only plausible ones (basically judges of different types, jurors, the government and tribunals). Because these exemplars share many features, it should be easy for subjects to list features of people who convict others and there should be substantial inter-subject consistency. In contrast, there are many plausible exemplars for the agent role of ACCUSE; basically any person can accuse someone else. On the other hand, the people who commonly accuse others seem to share some attributes that are relevant to the role. Thus it should be more difficult for people to list features for “someone who convicts others”, but still be easier than for many other roles (e.g. the agent role of SIT). A similar observation has been made in the realm of object concepts. For example, in McRae et al. (1997), subjects
found it easy to produce features for basic level concepts such as DOG, presumably because the features of individual dogs are densely correlated and/or overlap substantially. On the other hand, subjects found it difficult to produce coherent feature lists for heterogeneous, ambiguously bounded superordinate concepts such as FURNITURE. In summary, it was predicted that there would be variability among roles in terms of the consistency of listed features across subjects.

Method

Subjects. Thirty-two native English-speaking psychology undergraduates from the University of Western Ontario participated for course credit.

Materials. The stimuli consisted of 20 transitive verbs. These verbs were chosen because the experimenters intuitively felt they took animate agents and patients, thus holding animacy constant. The verbs also have reasonably restricted agent and patient roles, thus increasing the probability that subjects would find this to be a coherent task; that is, verbs like SIT were not included. In the two test packages, subjects were asked to list agent features for 10 of the verbs (e.g. “someone who frightens people”) and patient features for the other 10 (e.g. “someone who is arrested”). No subject listed features for the agent and patient roles of the same verb. Four verbs, each followed by 10 blank lines, were presented on each page. The agent roles were presented on the left side and the patient roles on the right. Packages were constructed to ensure that the subjects did not list features for the same role of two semantically similar verbs. For example, if a test package contained “someone who lectures people”, it contained “someone who is taught” rather than “someone who teaches people”, under the assumption that the agent concepts of LECTURE and TEACH are similar.

Procedure. Subjects were run in small groups in 45 minute sessions. Instructions were read and then left visible on a projector screen for the duration of the experiment. The instructions are included below because this is the first study in which subjects have been asked to list features for thematic roles. They were as follows.

Most verbs describe actions (e.g. hit, run). Much of the time, in an event described by a verb, someone is doing something to someone else. For many events (verbs), a variety of people or things can be involved. However, we are interested in the fact that there are certain types of people that are commonly associated with specific events. For example, certain people (e.g. hypnotist or Freudian psychologist) typically hypnotize others (e.g. their patient or a volunteer from the crowd). The types of people that commonly participate in
an action or event are often similar in some respects in that they tend to share certain features. For example, some features that might be shared by the kind of person who typically carries people are: [example here]. In this experiment, you will be asked to list features like those shown above. In addition to listing features for the type of people who tend to perform the action, you will be asked to list features for those who tend to have the action done to them. For example, below are listed some features that are common to people who are carried by other people: [example here]. Note that we did not list names of the individuals that commonly participate in these actions (e.g. baby or invalid for someone who is carried). Rather, we listed features that these individuals tend to share. Please list as many features as you can in the 10 spaces provided. You may list more than 10 features if you can think of them. You will have two minutes per item. Please do not move on to the next item until you are instructed to do so. Please do not exceed the two minute time limit.

Results and Discussion

A separate file was created for each role to catalogue and organise the responses. Features were recorded with their production frequency, that is the number of subjects who listed the feature for that particular role (maximum of 16). When recording the features, care was taken to record synonyms under a single label. For example, the features \( K \) is large and \( K \) is big were recorded under a single label, \( K \) is big. Many subjects also listed single-word features (e.g. \( K \) wise, \( K \) brave, \( K \) big). These one-word features were given the prefix \( K \) is to ensure a consistent presentation in Experiment 2. Care was also taken to ensure that features with different meanings were recorded using distinct labels. For example, both \( K \) is guilty and \( K \) may be guilty, were listed for “someone who is accused”. These features were recorded separately because it was felt that their meanings differ in an important way. Decisions concerning feature meaning were made with the assistance of three naive colleagues.

Subjects had no trouble providing features for the roles. Individual subjects listed an average of 5.5 features per role (5.7 per agent role, 5.3 per patient role). There were 3530 total responses, 1838 for agent roles, 1692 for patient roles. Of the 3530 responses, 1573 different features were listed, 800 for the agent roles and 773 for the patient roles. Thus each feature was listed by an average of 2.2 subjects, 2.3 for agent roles and 2.2 for patient roles. Of the 1573 features, 46% (721) were listed by at least 2 of the 16 subjects and 28% (445) by at least 3. The corresponding numbers were 48% (383/800) and 30% (237/800) for the agent roles and 44% (338/773) and 27% (208/773) for the patient roles. Similar statistics for consistency of responses were found in a recent feature norming experiment in which common object concepts such as DOG and CHAIR served as stimuli (McRae et al., 1997). In McRae and co-workers’ study, 42% of features were listed by 2 or more subjects and 27% by 3 or more (out of a possible 30 subjects).
The main purpose of Experiment 1 was to create agent and patient feature-based role representations for Experiment 2B. These representations included the features listed by at least 3 of 16 subjects for a role. Table 1 provides a flavour for these representations by presenting the role features for FRIGHTEN (the entire set of norms can be obtained from K.M.). Note that none of the features for these 40 roles resembled the conceptual primitives of Schlesinger (1995), nor the proto-agent or proto-patient features of Dowty (1991), and that this was true for all 40 role concepts. Rather, like those in Table 1, the features were specific to common participants in the events to which the verbs referred.

To investigate the variability in the ease with which subjects could produce features for specific roles, between-subject response consistency was computed. Consistency was first measured in terms of the number of features per role representation (i.e. features listed by at least 3 of 16 subjects), which was non-significantly greater for agent ($\bar{x} = 11.9$, range = 8–18) than for patient roles ($\bar{x} = 10.4$, range = 6–15) [$t(19) = 1.66, P > 0.1$]. A second measure was computed by dividing the number of features listed by three or more subjects by the number of semantically distinct features listed for that role. Table 2 presents the 20 verbs in decreasing order of the percentage of features listed by at least three subjects, averaged across the two roles. Again, response consistency was non-significantly greater for agent ($\bar{x} = 30\%$, range = 20–47\%) than for patient roles ($\bar{x} = 28\%$, range = 14–41\%) [$t(19) = 1.18, P > 0.2$]. However, it is clear that some roles are more well-defined than others. For example, there is substantial consistency across subjects for the agent roles of RESCUE (47\%), ENTERTAIN (45\%) and CONVICT (41\%), and the patient roles of CONVICT (41\%) and TEACH (40\%). In contrast, the lack of response consistency in the patient roles of ACCUSE (14\%), SERVE (14\%) and TERRORISE (15\%) reflects the fact that basically anyone can be accused, served or terrorised. In

<table>
<thead>
<tr>
<th>Agent Feature</th>
<th>Production Frequency</th>
<th>Patient Feature</th>
<th>Production Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>is mean</td>
<td>10</td>
<td>is scared</td>
<td>10</td>
</tr>
<tr>
<td>is scary</td>
<td>10</td>
<td>is small</td>
<td>7</td>
</tr>
<tr>
<td>is ugly</td>
<td>8</td>
<td>is weak</td>
<td>7</td>
</tr>
<tr>
<td>is big</td>
<td>7</td>
<td>is helpless</td>
<td>4</td>
</tr>
<tr>
<td>is sadistic</td>
<td>6</td>
<td>is jumpy</td>
<td>4</td>
</tr>
<tr>
<td>has problems</td>
<td>4</td>
<td>is nervous</td>
<td>4</td>
</tr>
<tr>
<td>is insensitive</td>
<td>4</td>
<td>is not knowledgeable</td>
<td>4</td>
</tr>
<tr>
<td>is heartless</td>
<td>3</td>
<td>is insecure</td>
<td>3</td>
</tr>
<tr>
<td>is unfriendly</td>
<td>3</td>
<td>is in trouble</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>is shuddering</td>
<td>3</td>
</tr>
</tbody>
</table>
TABLE 2
Number (and Percentage) of Features Listed by at Least Three Subjects in Experiment 1

<table>
<thead>
<tr>
<th>Verb</th>
<th>Agent Role</th>
<th>Patient Role</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Features</td>
<td>Number Listed by ≥ 3 Subjects</td>
</tr>
<tr>
<td>convict</td>
<td>37</td>
<td>15 (41%)</td>
</tr>
<tr>
<td>teach</td>
<td>41</td>
<td>14 (34%)</td>
</tr>
<tr>
<td>rescue</td>
<td>38</td>
<td>18 (47%)</td>
</tr>
<tr>
<td>entertain</td>
<td>29</td>
<td>13 (45%)</td>
</tr>
<tr>
<td>fire</td>
<td>42</td>
<td>12 (29%)</td>
</tr>
<tr>
<td>cure</td>
<td>28</td>
<td>11 (39%)</td>
</tr>
<tr>
<td>punish</td>
<td>45</td>
<td>12 (27%)</td>
</tr>
<tr>
<td>hire</td>
<td>40</td>
<td>12 (30%)</td>
</tr>
<tr>
<td>evaluate</td>
<td>30</td>
<td>9 (30%)</td>
</tr>
<tr>
<td>arrest</td>
<td>46</td>
<td>15 (33%)</td>
</tr>
<tr>
<td>lecture</td>
<td>45</td>
<td>9 (20%)</td>
</tr>
<tr>
<td>frighten</td>
<td>39</td>
<td>10 (26%)</td>
</tr>
<tr>
<td>instruct</td>
<td>33</td>
<td>9 (27%)</td>
</tr>
<tr>
<td>terrorise</td>
<td>43</td>
<td>15 (35%)</td>
</tr>
<tr>
<td>investigate</td>
<td>47</td>
<td>11 (23%)</td>
</tr>
<tr>
<td>worship</td>
<td>40</td>
<td>9 (22%)</td>
</tr>
<tr>
<td>interview</td>
<td>46</td>
<td>10 (22%)</td>
</tr>
<tr>
<td>accuse</td>
<td>46</td>
<td>13 (28%)</td>
</tr>
<tr>
<td>serve</td>
<td>44</td>
<td>12 (27%)</td>
</tr>
<tr>
<td>interrogate</td>
<td>41</td>
<td>8 (20%)</td>
</tr>
</tbody>
</table>

summary, there were clear differences among verbs in terms of consistency of responses, but the agent and patient roles did not differ reliably.

EXPERIMENT 2

Experiment 2 investigated whether the role representations created in Experiment 1 capture some basic aspects of people’s knowledge about the roles that certain entities play in events, and whether this knowledge is reflected by a view of thematic roles as verb-specific computed prototypes. In Experiment 2A, role/filler typicality was normed for the agent and patient roles of the Experiment 1 verbs. In this task, subjects provided ratings for questions such as, “How common is it for a hostage to rescue someone?”, or “How common is it for a hostage to be rescued by someone?” A 7-point scale was used, where 1 corresponded to very uncommon and 7 to very common. In Experiment 2B, different subjects were presented with the role features from Experiment 1 but were not informed of their source, and were asked to rate the importance/centrality of these features for a specific nominal filler concept. Each noun used in Experiment 2B was either a good agent or a good patient of its corresponding verb, as established by Experiment 2A. For example, a subject might rate the importance of the features of the agent
We wish to thank Michael Spivey-Knowlton, Maryellen MacDonald, Mike Tanenhaus and Neal Pearlmutter, who worked together with the first author in the collection of these role/filler typicality norms. The mean featural importance rating was used as a measure of role/filler featural similarity. This measure was then used to predict role/filler typicality. The logic of Experiment 2 mirrored that of previous studies in which feature-based concept similarity has been used to predict typicality ratings (e.g. Malt & Smith, 1984; Rosch & Mervis, 1975). However, this is the first time that it has been used in the context of thematic roles.

EXPERIMENT 2A

There were two aims. The first was to demonstrate that thematic roles have internal structure by showing that some nominal concepts are rated as better fillers than others (see Schlesinger, 1995, for similar rating experiments). The second was to use the agenthood and patienthood ratings to select nominal concepts that were relatively polarised for Experiment 2B. An effort was made to design a questionnaire that would reflect people’s judgements about events in the world, rather than about language per se. Therefore, instead of using, for example, a sentence acceptability judgement or a fragment completion task, subjects were asked how common they perceive certain events to be.

Method

Subjects. One hundred and forty-eight native English-speaking psychology undergraduates from the University of Southern California volunteered to participate. Note that no subject was involved in more than one of the experiments reported in this paper.

Materials. The items were 96 verbs, each paired with 3–7 nouns. These verbs included the 20 from Experiment 1, as well as those used by McRae et al. (in prep.), Pearlmutter and MacDonald (1992) and Trueswell et al. (1994). Nouns that the experimenters intuitively felt were either good agents or patients were chosen for each verb, or were taken from previously conducted experiments. To encourage subjects to use the middle of the scale, a small number of filler nouns were added to many of the verbs. Because there were a large number of verbs, they were divided into two lists of 48 to decrease the number of ratings required of each subject. Furthermore, each subject rated the nouns for either agenthood or patienthood, but not both. Thus four lists were created.

We wish to thank Michael Spivey-Knowlton, Maryellen MacDonald, Mike Tanenhaus and Neal Pearlmutter, who worked together with the first author in the collection of these role/filler typicality norms.
Procedure. Questionnaires were administered in a paper and pen format. Subjects were asked to rate on a 7-point scale how common it is for some type of person or animal to play a specific role in an event. Agenthood questions resembled example (1a) and patienthood questions resembled example (1b):

1a. How common is it for a
   snake  _______
   nurse  _______
   monster _______
   baby   _______
   cat    _______
   to frighten someone/something?

1b. How common is it for a
   snake  _______
   nurse  _______
   monster _______
   baby   _______
   cat    _______
   to be frightened by someone/something?

Examples were provided before subjects began the task. No time limit was imposed.

Results and Discussion

A good agent and a good patient were chosen from the 20 verbs of Experiment 1 based on two considerations. First, they were chosen to be near the ends of the continuum to facilitate analyses of variance in Experiment 2B, in which it was tested whether the featural importance ratings distinguished between good agents and patients. Second, they were chosen so that there was variation in the agenthood and patienthood ratings to facilitate regression analyses in Experiment 2B that tested whether role/filler featural similarity predicts role/filler typicality. Agenthood and patienthood ratings for the 20 good agents did not overlap [agenthood: $\bar{x} = 6.4$, range = 4.5–7.0; patienthood: $\bar{x} = 2.5$, range = 1.3–4.1; $t(19) = 20.68$, $P < 0.0001$]. The same was true for the good patients [agenthood: $\bar{x} = 1.9$, range = 1.2–3.5; patienthood: $\bar{x} = 6.1$, range = 5.2–7.0; $t(19) = 23.26$, $P < 0.0001$]. The good agents and patients are listed in Table 3, together with some statistics from Experiment 2B that are described as part of that experiment.

Although the majority of items in Experiment 2A were relatively polarised, there was a definite internal structure to the roles, as prototype theory would predict. Table 4 shows some example ratings to make this fact
<table>
<thead>
<tr>
<th>Verb</th>
<th>Agent Role</th>
<th>Patient Role</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Good Agent</td>
<td>Good Patient</td>
</tr>
<tr>
<td>accuse</td>
<td>witness</td>
<td>3.2</td>
</tr>
<tr>
<td>arrest</td>
<td>cop</td>
<td>5.4</td>
</tr>
<tr>
<td>convict</td>
<td>juror</td>
<td>4.6</td>
</tr>
<tr>
<td>cure</td>
<td>doctor</td>
<td>6.0</td>
</tr>
<tr>
<td>entertain</td>
<td>stripper</td>
<td>4.4</td>
</tr>
<tr>
<td>evaluate</td>
<td>committee</td>
<td>4.8</td>
</tr>
<tr>
<td>fire</td>
<td>employer</td>
<td>4.3</td>
</tr>
<tr>
<td>frighten</td>
<td>monster</td>
<td>5.2</td>
</tr>
<tr>
<td>hire</td>
<td>boss</td>
<td>4.8</td>
</tr>
<tr>
<td>instruct</td>
<td>coach</td>
<td>5.4</td>
</tr>
<tr>
<td>interrogate</td>
<td>inspector</td>
<td>4.4</td>
</tr>
<tr>
<td>interview</td>
<td>reporter</td>
<td>5.4</td>
</tr>
<tr>
<td>investigate</td>
<td>auditor</td>
<td>4.8</td>
</tr>
<tr>
<td>lecture</td>
<td>policeman</td>
<td>4.9</td>
</tr>
<tr>
<td>punish</td>
<td>babysitter</td>
<td>2.9</td>
</tr>
<tr>
<td>rescue</td>
<td>knight</td>
<td>5.8</td>
</tr>
<tr>
<td>serve</td>
<td>waitress</td>
<td>5.0</td>
</tr>
<tr>
<td>teach</td>
<td>professor</td>
<td>5.1</td>
</tr>
<tr>
<td>terrorise</td>
<td>pirate</td>
<td>4.1</td>
</tr>
<tr>
<td>worship</td>
<td>priest</td>
<td>3.9</td>
</tr>
</tbody>
</table>

* A non-significant difference.
TABLE 4
Some Example Role/Filler Typicality Ratings

<table>
<thead>
<tr>
<th>Verb</th>
<th>Noun Filler</th>
<th>Agenthood Rating</th>
<th>Patienthood Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>frighten</td>
<td>monster</td>
<td>6.4</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>snake</td>
<td>6.2</td>
<td>3.9</td>
</tr>
<tr>
<td></td>
<td>nurse</td>
<td>3.8</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>cat</td>
<td>3.1</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>baby</td>
<td>2.0</td>
<td>5.5</td>
</tr>
<tr>
<td>hire</td>
<td>manager</td>
<td>6.7</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>boss</td>
<td>6.7</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>chairman</td>
<td>4.5</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>employee</td>
<td>1.9</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>applicant</td>
<td>1.3</td>
<td>5.6</td>
</tr>
</tbody>
</table>

apparent. Presumably, a judiciously chosen large sample of nouns would result in a more complete and relatively smooth gradient, although this was not fully tested in this experiment. It is concluded that the conceptual categories corresponding to agent and patient roles are not classical in nature, as theories that incorporate selectional restrictions might suggest.

EXPERIMENT 2B

The purpose was to test whether the internal structure of agent and patient roles could be captured by a measure of the similarity between role concepts and nominal fillers, henceforth referred to as role/filler featural similarity. The experiment used a good agent and patient for each verb selected on the basis of Experiment 2A role/filler typicality.

Method

Subjects. Forty native English-speaking psychology undergraduates from the University of Western Ontario participated for course credit.

Materials. Stimuli were constructed by crossing the good agents and patients from Experiment 2A with the role features of the 20 verbs from Experiment 1 (2 roles × 2 fillers × 20 verbs = 80 items). Four lists were created, each containing 20 items. Each verb was represented once on each list. For example, agent features for RESCUE appeared in one list with KNIGHT and another with HOSTAGE. The other two lists paired KNIGHT and HOSTAGE with the patient features for RESCUE. A 7-point scale was presented at the top of each test page, ranging from “not at all important/central” to “very important/central”. Four items were presented on each page and the pages were randomly ordered.
Procedure. Subjects were run in small groups. Each subject received one list. The instructions were read to subjects and remained on a projector screen throughout. Subjects were asked to rate on a 7-point scale how important/central each feature was to their concept of the noun in question. The source of the features was not divulged. No time limit was imposed. On average, the study took about 15 min to complete.

Results and Discussion. Mean feature centrality rating across subjects was computed for each feature and an example (FRIGHTEN) is presented in Table 5. Role/filler featural similarity was computed as the mean feature centrality rating for each filler and is presented in Table 3 above.

The first analyses were designed to demonstrate that role/filler featural similarity captured basic aspects of the relationship between the roles and fillers. They tested whether role/filler featural similarity is greater when the filler matches the role. For example, it was predicted that role/filler featural similarity would be greater for MONSTER/FRIGHTEN’s agent role features than for BABY/FRIGHTEN’s agent role features, but less when the features come from FRIGHTEN’s patient role. Analyses of variance were conducted using role (agent vs patient features) and congruency (whether the filler matches the role) as independent variables and role/filler featural similarity as the dependent variable. As predicted, role/filler featural similarity was significantly higher for congruent ($\bar{x} = 4.5, SE = 0.1$) than for incongruent fillers ($\bar{x} = 2.7, SE = 0.1$) [$F_1(1,38) = 166.94, P < 0.0001; F_2(1,38) = 100.36, P < 0.0001$]. Interestingly, role and congruency interacted [$F_1(1,38) = 9.73, P < 0.004; F_2(1,38) = 8.11, P < 0.008$]. Planned comparisons showed that the source of the interaction was a greater difference between congruent and incongruent fillers for agent roles [congruent: $\bar{x} = 4.7, SE = 0.1$; incongruent: $\bar{x} = 2.4, SE = 0.2$; $F_1(1,38) = 128.75, P < 0.0001; F_2(1,38) = 36.63, P < 0.0001$] than for patient roles [congruent: $\bar{x} = 4.3, SE = 0.1$; incongruent: $\bar{x} = 2.9, SE = 0.2$; $F_1(1,38) = 34.58, P < 0.0001; F_2(1,38) = 20.41, P < 0.0001$]. Differences between roles are discussed below. Finally, there was no main effect of role ($F < 1$ in both analyses).

To complement the analyses of variance, regression analyses were conducted in which role/filler featural similarity was used to predict role/filler typicality that had been measured in Experiment 2A. Regressions were conducted on the set of data as a whole, and on the two roles separately. First, collapsed across the two roles, role/filler featural similarity predicted 59% of the variance in role/filler typicality [$r = 0.77; F(1,78) = 111.90, P < 0.0001$]. Figure 1 graphically depicts this strong correlation. The second analysis investigated the typicality gradient of the agent role. Role/filler featural similarity between the agent concepts and the good agents and patients predicted 75% of the variance in role/filler typicality [$r = 0.87; F(1,38) = 114.32, P < 0.0001$]. For the patient role, role/filler featural
similarity predicted 39% of the variance in role/filler typicality \( [r = 0.62; F(1,38) = 23.84, P < 0.0001] \). A test of the difference between correlations revealed that role/filler featural similarity better captured role/filler typicality for the agent role than for the patient role \( (z = 2.53, P < 0.02) \).

To explore potential differences between agent and patient roles further, additional analyses tested for a congruency effect within each role of each verb; that is, paired \( t \)-tests determined whether role/filler featural similarity differed for congruent versus incongruent fillers for each role (rather than overall, as was investigated above). These analyses revealed that 34 of the 40 roles followed the general pattern; that is, role/filler featural similarity was higher for the congruent filler \( (P < 0.05) \). The results are presented in Table 3. Asterisks designate the six non-significant differences, those being the agent role of ACCUSE and the patient roles of CURE, HIRE, INTERVIEW, TEACH and WORSHIP. Thus this set of analyses again showed that role/filler featural similarity better captured role/filler typicality for the agent roles than for the patient roles. Insight into this result might be gained by considering the patient roles of the five verbs for which role/filler featural similarity did not adequately capture role/filler typicality. The main

### Table 5

<table>
<thead>
<tr>
<th>Role</th>
<th>Feature</th>
<th>Importance Rating</th>
<th>Importance Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Good Agent</td>
<td>Good Patient</td>
</tr>
<tr>
<td>Agent</td>
<td>monster</td>
<td>baby</td>
<td></td>
</tr>
<tr>
<td>is mean</td>
<td>5.5</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>is scary</td>
<td>6.3</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>is ugly</td>
<td>5.7</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>is big</td>
<td>5.9</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>is sadistic</td>
<td>3.7</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>has problems</td>
<td>3.2</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>is insensitive</td>
<td>5.0</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>is heartless</td>
<td>5.2</td>
<td>4.6</td>
<td></td>
</tr>
<tr>
<td>is unfriendly</td>
<td>6.1</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Patient</td>
<td>monster</td>
<td>baby</td>
<td></td>
</tr>
<tr>
<td>is scared</td>
<td>2.5</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>is small</td>
<td>2.3</td>
<td>5.7</td>
<td></td>
</tr>
<tr>
<td>is weak</td>
<td>2.2</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td>is helpless</td>
<td>1.5</td>
<td>5.6</td>
<td></td>
</tr>
<tr>
<td>is jumpy</td>
<td>2.9</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>is nervous</td>
<td>1.3</td>
<td>3.1</td>
<td></td>
</tr>
<tr>
<td>is not knowledgeable</td>
<td>1.2</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td>is insecure</td>
<td>2.0</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>is in trouble</td>
<td>2.3</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>is shuddering</td>
<td>2.0</td>
<td>2.2</td>
<td></td>
</tr>
</tbody>
</table>
factor appears to be that these role categories sensibly admit many exemplars; that is, almost everyone has been cured, hired, interviewed and taught (although this does not extend as well to being worshipped). Thus, in these cases, the good agents actually fit the patient features quite well, whereas this was not the case for more restricted patient roles like that of ARREST.

Further support for the possibility that the patient roles generally admit a wider range of fillers was obtained by considering the role/filler typicality of the incongruent fillers (role/filler featural similarity was not included in these analyses). A $t$-test was conducted to contrast the roles in terms of the role/filler typicality of their incongruent fillers. Note that these ratings should be low for both roles precisely because the fillers were incongruent. Although role/filler typicality was relatively low in both cases, it was significantly higher for the good agents in the patient role ($\bar{x} = 2.5$, SE = 0.2)
than for the good patients in the agent role ($\bar{x} = 1.9$, SE = 0.2) [$t(38) = 2.46$, $P < 0.02$].

In the final analysis, the incongruent fillers were ranked in descending order of role/filler typicality. Seven of the top 10 and 15 of the top 19 ratings for incongruent fillers came from patient roles, again suggesting that the patient roles admit a greater range of exemplars.

These results cohere with the consistency analyses in Experiment 1, in which it was found that the features listed for patient roles were less consistent across subjects than for agent roles, although the difference did not reach significance. Thus it would seem that the patient roles of the verbs used in Experiments 1 and 2 are less well-defined than the agent roles. Presumably, this is a function of the distribution of the exemplars in the role concepts; a set of common, featurally similar exemplars would result in a distinct computed prototype. A role might not have a distinct computed prototype if the set of common fillers have few featural regularities among them, or if there is no set of exemplars that stand out in terms of the frequency of playing the role. The latter appears to be the important factor in the five patient roles mentioned above.

Previous researchers have noted that verbs vary in terms of the role information that they carry with them, particularly cross-linguistically. For example, Nagy and Gentner (1990) pointed out that English verbs frequently incorporate information about the instrument used to perform the specified action, but rarely incorporate information about patient features. In contrast, Talmey (1975) has noted that German differs in that there is a greater tendency for verbs to specify patient features. The differences between agent and patient role concepts presented above indicate a tendency in English for verbs to carry more information about their agents than their patients. It is possible, however, that the tendencies in English might vary systematically. For instance, the verbs used in Experiments 1 and 2 were chosen to favour animate patients and one fact about the world is that many things that happen to people can happen to many types of people; thus producing indistinct patient role concepts in these verbs. However, it is easy to think of verbs for which the patient role has a distinct prototype, particularly those that favour inanimate patients. For example, only air is breathed (although one can breathe pure oxygen or other components of air) and only a few things are raked, such as lawns, gardens and flower beds. Thus it is possible that the distinctiveness of role prototypes varies across verbs according to such general semantic principles.

In summary, in accord with many studies of natural concepts (e.g. Rosch & Mervis, 1975), thematic role concepts exhibit a typicality gradient, with category/exemplar (role/filler) typicality being predictable from a measure of featural similarity. Thus these experiments taken together demonstrate that certain aspects of event concepts act like nominal concepts. Due to the
focus on nominal concepts in the concepts literature, this is the first time that this link has been made.

**EXPERIMENT 3**

Experiments 1 and 2 demonstrated that the notion of feature-based role concepts can be used to capture important aspects of people’s knowledge about events and their common participants. Experiment 3 tested the claim that this knowledge should be viewed as a primary component of thematic roles by showing that it is computed and used quickly in on-line language comprehension. Although a number of linguists and psycholinguists have recognised that the knowledge studied in Experiments 1 and 2 is important to language processing, they have retained the distinction between it and grammatical knowledge, of which they assume thematic roles are a part (Schlesinger, 1995). For example, Schlesinger was explicit in stating that conceptual primitives relevant to thematic roles such as \textit{cause} and \textit{change} are part of the lexical representation of the verb, but features of the type studied above become active only through a delayed inference process (i.e. world knowledge is stored somewhere else in the system and it takes time to access it). Likewise, in the garden path theory of Frazier and colleagues (e.g. Frazier, 1987; Rayner, Carlson, & Frazier, 1983), this information is part of the “thematic processor” and influences language comprehension only after a delay. In direct contrast, the present claim is that a verb automatically and immediately makes available thematic role concepts. Furthermore, these role concepts immediately influence on-line comprehension and production, primarily because of the constraint-based nature of sentence processing. In the remainder of this introduction, some recent evidence is presented suggesting that this is so, and an on-line reading experiment is introduced that further tested these claims.

**Immediately Available**

McRae, Ferretti and Airey (1996) presented three automatic priming studies to demonstrate that verb-specific thematic role concepts become available immediately and automatically. In the first experiment, verbs (e.g. CONVICTING) were used to prime their typical agents and patients (e.g. JUROR and CRIMINAL). Verbs such as LOVING that tend to be used with animate agents and patients were used as primes in the unrelated control condition. A significant overall priming effect was obtained in a semantic decision task when primes and targets were presented visually with a stimulus onset asynchrony of 250 msec. When analysed by role, however, the priming effect was significant for patients (33 msec), but not for agents (9 msec). Presumably, this difference was due to the fact that in English, agents usually precede the verb and patients follow it, whereas in the
priming experiment, both agents and patients directly followed their verb prime. Therefore, in a subsequent experiment, the same verbs were placed in short sentence fragments so as to have both agents and patients follow the verb in a natural manner (e.g. “The man convicted the” or “The man was convicted by the”). When subjects heard a fragment and then cross-modally named a good agent or patient, priming was found for both agents (19 msec) and patients (25 msec). Finally, McRae et al. tested for automatic activation of role features by priming adjectival patient features (e.g. gullible) with verbs presented in isolation (e.g. TRICKING). A significant 33 msec priming effect was found in a lexical decision task when verb primes and feature targets were presented visually with a stimulus onset asynchrony of 250 msec. These three experiments demonstrate that verb-specific thematic role knowledge is activated immediately and automatically, and that this activation is influenced by structural factors.

Immediately Used

Given that this information is available, the next issue is how quickly it influences sentence interpretation. On-line experiments designed to investigate this question actually test two related hypotheses simultaneously. First, if this type of knowledge is used immediately to aid the assignment of noun fillers to their roles, then it is appropriate to view it as thematic role knowledge. This conclusion would contrast with Schlesinger (1995) and others, who claim that information about events and their participants is a product of inferencing and comes into play only at a later stage. Second, if thematic role concepts immediately influence comprehension, then this stands as evidence against sentence processing theories that highlight a front-end syntactic module, the most prominent of which is the garden path theory of Frazier and colleagues (Ferreira & Clifton, 1986; Frazier, 1987; Rayner et al., 1983). In principle, a theory of thematic roles as verb-specific concepts can be integrated with either the constraint-based theory of sentence processing (e.g. MacDonald et al., 1994), or with the garden path theory as part of the thematic processor (Rayner et al., 1983). However, given our claims concerning the automaticity and speed with which this conceptual knowledge is assumed to influence comprehension, it is a better fit with the constraint-based theory in which all information is computed as quickly as possible and used as soon as it is available, than with the garden path theory, in which syntactic processing is given temporal priority over conceptual processing.

The on-line process of designating local NPs to a verb’s thematic roles is known as thematic assignment. We assume that assignment is based on a number of cues, including structural regularities, conceptual knowledge and discourse context. The present work focused on conceptual knowledge,
remaining agnostic as to the form in which structural regularities are represented and implemented (for contrasting views on this issue, see Frazier, 1995; MacDonald et al., 1994), as well as the precise mechanism that allows for the interaction between structural and conceptual cues (see Gropen et al., 1991, for one account). Thematic fit was operationalised in Experiment 3 as role/filler typicality (as it was in McRae et al., in prep.; Pearlmutter & MacDonald, 1992; Tabossi, Spivey-Knowlton, McRae, & Tanenhaus, 1994), which presumably reflects role/filler featural similarity, as shown in Experiment 2. Thus it was assumed that thematic assignment depends on the on-line activation of a role concept, which is akin to activating a set of features that may or may not match a single lexical concept, and may overlap to varying degrees with a number of lexical concepts. Thematic assignment is then a form of speeded categorisation, in which a candidate noun filler concept can be compared to a role concept, resulting in a judgement of varying certainty (Hampton, 1993; Smith, 1978). Note that this contrasts sharply with another possible view in which a list of words (i.e. possible fillers) is searched and assignment is based on whether or not the candidate filler is found (as in verification models of word recognition, such as Becker, 1980). Importantly, thematic fit is assumed to be a primary determinant of assignment, particularly when syntactic cues underdetermine role assignments, as in cases of local structural ambiguity.

McRae et al. (in prep.), Pearlmutter and MacDonald (1992) and Tabossi et al. (1994) have demonstrated verb-specific effects of thematic fit on the resolution of the main clause/reduced relative clause ambiguity. McRae et al. described below because it is the clearest demonstration of immediate effects and is most related to the present work. McRae et al. contrasted people’s ability to resolve a main clause/reduced relative clause ambiguity when the initial NP, which was always animate, was either a good patient/poor agent such as CUSTOMER (2a) or a good agent/poor patient such as WAITRESS (2b), as established by role/filler typicality norms:

2a. The customer (who was) served by the trainee was displeased with his attitude.
2b. The waitress (who was) served by the trainee was displeased with his attitude.

In English, a sentence-initial “NP verbed” sequence usually signals a main clause construction (Tabossi et al., 1994). Furthermore, sentence-initial “NP verbed” sequences typically translate to “agent verbed” (Bever, 1970). Because sentences like (2a) and (2b) violate this regularity, correctly interpreting them requires the reader or listener to make the less preferred assignments. In a self-paced reading experiment, McRae et al. found that readers made immediate use of thematic fit information in that thematic fit and reduction interacted in the verb + by region, where reduction refers to
the difference between unreduced (“who was” present) versus reduced (“who was” absent). Furthermore, sentences like (2a) that began with a good patient were resolved more quickly than ones like (2b) that began with a good agent. In addition, ambiguity resolution was simulated by a competition model in which thematic fit played a major role and was not delayed. These experiments indicate that the knowledge explored in Experiments 1 and 2 is best conceived of as thematic role knowledge and that this information is computed and used immediately, contrary to researchers who make a temporal distinction between structural and conceptual processing, such as Schlesinger (1995) and Frazier (1987).

Biasing On-line with Features

The goal of Experiment 3 was to add to both the thematic role and structural ambiguity resolution literatures by determining whether role feature information could modulate the processing difficulty associated with reading temporarily ambiguous reduced relative clauses. Just as features can be reported linguistically as adjectives, they can be presented in sentences as adjectives. Thus to bias the initial NP towards being a good agent or patient, features were used that were associated with verb-specific agent and patient roles. The items had the following structure (the complete set is listed in the Appendix).

3a. The shrewd heartless gambler (who was) manipulated by the dealer had bid more than he could afford to lose.
3b. The young naive gambler (who was) manipulated by the dealer had bid more than he could afford to lose.

For the thematic fit manipulation to be effective in this case, the reader needs to perform a number of computations. First, she must interpret the complex initial NP (adjective adjective noun). If the entire NP is not understood, no effect of thematic fit can arise because the base noun (“gambler”) is identical in both conditions. Murphy (1990) and Smith et al. (1988) have discussed some ways in which adjective–noun NPs might be computed on the fly. In addition, Murphy has shown that sensible NPs consisting of a noun modified by predicting adjectives are understood quite quickly and easily, and all NPs used in Experiment 3 were of this type. (A predicating adjective is one that can be used in a “The noun is adjective” construction, such as “The gambler is young”). The reader must then compute the fit between the complex NP and the agent and patient roles of the initial verb. Thus a substantial amount of semantic processing must be completed quickly if the manipulation is to have an early influence on ambiguity resolution.
There are a number of constraints other than thematic fit at work in these sentences, each of which was held constant within a pair like (3a) and (3b). Three of the constraints were available at the past participle (“manipulated”). The strongest was a general structural bias in favour of a main clause; in a recent analysis of text, it was found that a sentence-initial sequence of “NP verbed” begins a main clause 92% of the time (Tabossi et al., 1994). In addition, a referential bias worked against finding an immediate effect of thematic fit. Referential theory suggests that the more that a noun is modified, the less felicitous it is to further modify it with a relative clause (Altmann & Steedman, 1988). This notion is pertinent to Experiment 3, in which two adjectives served as pre-nominal modifiers, then a post-nominal (reduced) relative clause followed. Third, the initial nouns were either relatively balanced or favoured a main clause reading, depending on one’s perspective. According to a role/filler typicality rating norming study reported below, the initial nouns were good agents and patients, although they were better patients than agents. However, according to a theory in which thematic roles contain only a few selectional restrictions, the nouns would be agent-biased because all of them were animate and paired with verbs that prefer animate agents.

In addition to the constraints available at the verb, post-verbal constraints combined to provide support for a reduced relative reading. The post-verbal by-phrase probabilistically cued readers that they were in a reduced relative clause (MacDonald, 1994). However, a by-phrase can also indicate manner, location, time or amount, but these interpretations were largely blocked by two additional factors. First, the experimenters judged the verbs (actually, past participles) as transitive-biased (i.e. are usually used with a direct object), thus increasing the probability that the by-phrase contained its agent. Second, the noun in the by-phrase was a good agent, but a very poor manner, location, time and amount (e.g. “dealer” in 2a and 2b), further biasing the reader towards a reduced relative interpretation.

It was predicted that sentences like (3a), which used patient-biasing features, would be easier to read than those like (3b), which used agent-biasing features. As a slight qualification, it should also be noted that because of the subtlety of the featural manipulation, the complexity of the initial NP, and the referential bias, it seemed unlikely that an immediate effect of thematic fit would be found. This is important in terms of the way in which Experiment 3 bears on the debate between constraint-based and garden path theories. The critical difference between the empirical predictions of these theories hinges on whether a non-structural or lexically specific constraint influences interpretation at the earliest possible moment. Because early effects do not categorically distinguish between the theories, it was expected that this experiment would not provide definitive evidence, but instead would contribute
to the debate by showing that this information can be computed and used rapidly.

Method

**Subjects.** Forty-eight native English-speaking psychology undergraduates from the University of Western Ontario received course credit for their participation.

**Materials.** Twenty noun–verb combinations were created with the constraint that the noun was both a good agent and patient of the verb (e.g. GAMBLER MANIPULATED). Note that we were able to use only a few of the verbs and features from Experiments 1 and 2 because of the difficulty in devising nouns for those verbs that both sounded natural and were approximately balanced in terms of their agenthood and patienthood. The newly created set of items was normed in two ways. First, 20 subjects provided role/filler typicality ratings for the nouns. Although the nouns were rated as good agents and patients, they were rated as better patients ($\bar{x} = 5.1$, range = 3.6–6.7) than agents ($\bar{x} = 4.4$, range = 2.7–6.2) for their respective verbs [$t(19) = 2.24, P < 0.04$].

Each of the 20 noun–verb combinations was combined with two pairs of adjectival features; one pair biased the noun towards being a good agent but a poor patient of the verb, and the other biased the noun towards being a good patient but a poor agent. Because we were unable to use the verbs from Experiments 1 and 2, the adjectives were selected on the basis of an independent norming study in which 30 subjects rated how common it is for a particular noun concept that is modified by two adjectives to fill either the agent or patient role of a particular verb, as in the following:

4. How common is it for a  
   friendly witty  
   shrewd heartless  
   gullible sensitive  
   mean talented  
   young naive  
   gambler to manipulate someone?

Subjects rated either the agenthood or patienthood of each item on a 7-point scale, where 1 corresponded to very uncommon and 7 to very common. Agent-biased NPs received significantly higher agenthood ($\bar{x} = 6.3$, range = 4.5–7.0) than patienthood ratings ($\bar{x} = 3.8$, range = 1.8–6.0) [$t(19) = 10.18, P < 0.0001$]. Similarly, patient-biased NPs received significantly higher patienthood ($\bar{x} = 5.3$, range = 2.8–5.8) than agenthood ratings ($\bar{x} = 2.0$, range = 1.1–4.5) [$t(19) = 14.30, P < 0.0001$].
Each of the resulting 20 noun–verb combinations was treated as an item (see Appendix). Four target sentences were constructed from each item by crossing featural bias (agent vs patient) and reduction (“who was” present vs absent). A second sentence was added to each target to increase the meaningfulness of the narrative. The following is an example.

5a. The shrewd heartless gambler who was manipulated by the dealer had bid a lot more money than he could afford to lose. Fortunately for him, he won.

5b. The young naive gambler who was manipulated by the dealer had bid a lot more money than he could afford to lose. Fortunately for him, he won.

The four versions were placed in separate lists to ensure that no subject saw any target verb or NP more than once. All of the noun–verb combinations were contained in each list, so that each contained five agent-biased unreduced, five agent-biased reduced, five patient-biased unreduced and five patient-biased reduced sentences. To dampen order effects, two versions of each of the four lists were formed by reversing the order of the target items. Sixty sentence pairs not containing a relative clause were included so that targets never occurred consecutively nor appeared first in a list. Fourteen additional sentence pairs without relative clauses were used for practice. Adjectives were included in many of the filler sentences.

Procedure. Sentences were displayed on a 16 inch Sony Trinitron monitor controlled by a Macintosh LCIII. They were presented using PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993) in a two-word-at-a-time moving window format. Two-word-at-a-time presentation, in contrast to one-word-at-a-time, was used because it more accurately mimics key elements of reading reduced relative clauses that contain a post-verbal by-phrase (Burgess, 1991). Thus sentence pairs were initially presented on the screen with each non-space character replaced by a dash. Subjects pressed a button to reveal the first two words of the sentence. Each subsequent button press revealed the next two words and replaced the previous two with dashes. Subjects read each sentence pair in this manner and then answered a “yes/no” comprehension question. Testing sessions began with 14 practice items. Subjects then read the remaining 80 experimental trials, taking a break after every 20 trials. They were instructed to read at a pace that resembled how they would typically read a magazine or newspaper. A testing session lasted approximately 30 min. Two-word-at-a-time reading latencies were recorded with millisecond accuracy via a CMU button box and measured as the time interval between successive button presses.
Results. Reading latencies were analysed based on the word pairs presented to subjects. The three critical regions are shown below:

6. The young naive gambler / manipulated by / the dealer / had bid / verb + by agent NP main verb

more than he could afford to lose.

Analyses of variance were conducted for the three regions using two-word reading latency as the dependent variable. There were two factors, each with two levels: featural bias (agent vs patient) and reduction (unreduced vs reduced). Both featural bias and reduction were within subjects and items. Planned comparisons were conducted to investigate the reduction effect separately for agent-biased and patient-biased items in all three regions. Reduction effects for the three critical regions are presented graphically in Fig. 2.

Verb + by. Featural bias showed little or no influence on reading latencies in this region. Bias and reduction did not interact ($F < 1$ in both analyses). Planned comparisons revealed that subjects experienced difficulty with both agent-biased [451 vs 512 msec: $F_1(1,47) = 15.04$, $P < 0.0004$; $F_2(1,19) = 18.95$, $P < 0.0004$] and patient-biased sentences [464 vs 530 msec: $F_1(1,47) = 17.42$, $P < 0.001$; $F_2(1,19) = 22.12$, $P < 0.0003$]. Overall, reading latencies for patient-biased sentences were a non-significant 15 msec longer [$F_1(1,47) = 1.92$, $P > 0.1$; $F_2 < 1$]. Finally, the average reading latency for unreduced sentences was 63 msec shorter than their reduced counterparts [$F_1(1,47) = 24.09$, $P < 0.0001$; $F_2(1,19) = 13.86$, $P < 0.002$].

Agent NP. Featural bias influenced reading latencies in the “determiner + noun” region (the actual agent of the past participle). The interaction between bias and reduction failed to reach significance [$F_1(1,47) = 2.52$, $P > 0.1$; $F_2(1,19) = 2.09$, $P > 0.1$]. However, Fig. 2 and planned comparisons show that subjects had no difficulty with patient-biased reduced relatives (478 vs 479 msec: $F < 1$ in both analyses), but reading difficulty was evident with agent-biased reduced relatives [458 vs 500 msec: $F_1(1,47) = 5.43$, $P < 0.03$; $F_2(1,19) = 4.42$, $P < 0.05$]. Finally, there was no main effect of bias (1 msec effect: $F < 1$ in both analyses) nor of reduction [22 msec effect: $F_1(1,47) = 2.11$, $P > 0.1$; $F_2(1,19) = 1.32$, $P > 0.2$].

Main Verb. In this region, reading difficulty for a reduced relative clause had been basically eliminated for both types of sentences; however, patient-biased sentences were read faster overall. Bias and reduction did not interact ($F < 1$ in both analyses). Planned comparisons revealed that
subjects had little difficulty reading reduced relatives, regardless of whether the features were patient-biasing (458 vs 469 msec) or agent-biasing (483 vs 492 msec) ($F < 1$ in all analyses). There was, however, a 23 msec main effect of featural bias that was significant by items [$F_2(1,19) = 6.03, P < 0.03$] and marginal by subjects [$F_1(1,47) = 3.50, P < 0.07$]. Finally, the 11 msec reduction effect was non-significant ($F < 1$ in both analyses).

**Discussion**

Experiment 3 showed that manipulating the fit between readers’ role concepts and complex NPs rapidly influenced their ability to resolve a main clause/reduced relative clause ambiguity. When the initial noun was preceded by patient-biasing features, readers used the thematic information to minimise the difficulty associated with a sentence-initial reduced relative clause. This experiment adds to those of McRae et al. (1996; in prep.), Pearlmutter and MacDonald (1992) and Tabossi et al. (1994) in
Note that Ferreira and Clifton (1986) failed to find an effect of initial NP animacy. However, Trueswell et al. (1994) demonstrated that the knowledge studied in Experiments 1 and 2 is computed and used quickly in on-line language processing, thus suggesting that role concepts are linguistically relevant.

The influence of featural bias found in Experiment 3 was not as strong as the effects of McRae et al. (in prep.) nor the animacy effects of Trueswell et al. (1994). However, the present manipulation was somewhat weaker and more complex. Trueswell et al. contrasted inanimates that were virtually impossible as agents of the verb (e.g. “The bricks lifted”) with animate good agents (e.g. “The workers lifted”). McRae et al., as well as Pearlmutter and MacDonald (1992), contrasted strongly agent-biased animates with strongly patient-biased ones. By using different nouns, the entire nominal concept was manipulated, although there was presumably some conceptual overlap between good agent and patient. This is equivalent to manipulating many of the features of the entity, as well as making salient the relevant ones (e.g. “The cop/crook arrested”). In contrast, in Experiment 3, only two features were manipulated at a time (plus perhaps features that were strongly correlated with them), with the initial noun remaining constant for the agent- versus patient-biased sentences. Thus even though thematic fit influenced ambiguity resolution quite rapidly (at the “determiner + noun” region), Experiment 3 cannot be taken as definitive evidence against the garden path model of Frazier and colleagues (Frazier, 1987; Rayner et al., 1983). However, it does provide yet another example of a rapid effect of non-structural information on on-line language comprehension and seems better suited to the principles of constraint-based models (MacDonald et al., 1994).

We have claimed that there is a verb-specific component to thematic role knowledge in addition to their verb-general aspects that are typically studied. An interesting question concerns the degree to which the manipulation in Experiment 3 should be considered as reflecting verb-specific role concepts. The answer depends on the rival theory. Given a theory in which the conceptual content of a thematic role is limited to syntactically relevant features such as animacy (e.g. Caplan et al., 1994), the answer is unequivocally “yes”. Because animacy was equated in Experiment 3, the results demand an explanation in which additional information is being computed and used rapidly. Furthermore, McRae et al. (in prep.) have found immediate effects of thematic fit using solely animate nouns.

On the other hand, a well-articulated theory that is more similar to the present one is that of Dowty (1991), which is weakly verb-specific. Dowty postulates five proto-agent and proto-patient features. Although the

\footnote{Note that Ferreira and Clifton (1986) failed to find an effect of initial NP animacy. However, Trueswell et al. (1994) corrected a number of problems with Ferreira and Clifton’s experiment and found a strong and immediate effect of animacy when those problems were alleviated.}
features are general in nature, such as \{volitional involvement in the event or state\} and \{undergoes change of state\}, the theory is verb-specific in that a verb’s role might consist of any number of these features. It is weakly verb-specific because many verbs within broad semantic classes would be expected to have identical role representations. In contrast, the present theory incorporates features that more subtly differentiate among verbs. The pertinent question here is whether the features used in Experiment 3 biased readers generally towards agenthood or patienthood (i.e., in a Dowty way) or towards agent and patient role concepts of the specific verb. We attempted to answer this question by conducting role/filler typicality rating experiments in which the verb was omitted. In these experiments, subjects provided ratings for questions such as “How common is it for a shrewd heartless person to have something done to them?” or “How common is it for a shrewd heartless person to have something done to them?” Both the agent-biased [agenthood: $\bar{x} = 4.9$, range = 2.2–6.6; patienthood: $\bar{x} = 3.8$, range = 2.7–5.3; $t(19) = 3.28$, $P < 0.004$] and patient-biased [agenthood: $\bar{x} = 2.9$, range = 1.0–6.8; patienthood: $\bar{x} = 4.9$, range = 2.6–6.1; $t(19) = 4.39$, $P < 0.0004$] adjective pairs received significantly higher ratings for their congruent general role. Similar results were obtained when the noun was included, as in “How common is it for a shrewd heartless gambler to have something done to them?”: agent-biased [agenthood: $\bar{x} = 4.6$, range = 2.2–6.4; patienthood: $\bar{x} = 3.8$, range = 2.5–5.3; $t(19) = 2.37$, $P < 0.03$]; patient-biased [agenthood: $\bar{x} = 2.6$, range = 1.1–6.3; patienthood: $\bar{x} = 4.1$, range = 2.4–5.7; $t(19) = 5.27$, $P < 0.0001$]. Finally, the three types of ratings were significantly correlated: verb-specific features vs verb-general features ($r = 0.77$, $P < 0.0001$ for agent-biased; $r = 0.36$, $P < 0.03$ for patient-biased); verb-specific versus verb-general features plus noun ($r = 0.72$, $P < 0.0001$ for agent-biased; $r = 0.43$, $P < 0.01$ for patient-biased); verb-general features only versus verb-general features plus noun ($r = 0.91$, $P < 0.0001$ for agent-biased; $r = 0.18$, $P > 0.2$ for patient-biased). Thus the stimuli did not differ substantially on the basis of these ratings.

One method that might empirically distinguish between Dowty-type effects and those of the present theory is to shuffle the adjectives and/or nouns among the sentences. If the information that gave rise to the effect of thematic fit in Experiment 3 was indeed at the level of Dowty’s features, then the precise pairings of the nouns, adjectives and verbs were irrelevant. Therefore, the same results should be obtained if the adjectives were randomly recombined with different nouns and the noun–verb combinations were retained. The results should also be obtained if the NPs were kept together and recombined with different verbs. Although these experiments have not been conducted, it is highly unlikely that effects of thematic fit would be found with random recombinations, even if care was taken to avoid nonsensical sentences.
GENERAL DISCUSSION

The experiments presented above, particularly when considered in conjunction with those of McRae et al. (1996; in prep.), Pearlmutter and MacDonald (1992) and Tabossi et al. (1994), outline and provide evidence for a novel view of thematic roles; namely, the notion of thematic roles as verb-specific concepts. In Experiment 1, subjects listed features for roles such as “someone who is convicted” and these featural descriptions were used to create approximations to computed role concepts. Experiment 1 also showed that some roles are better defined than others, presumably because there are important differences in the regularities among the exemplars that form their bases. In Experiment 2A, subjects’ ratings of role/filler typicality showed that role concepts have internal structure, as predicted when roles are viewed as computed prototypes. Experiment 2B demonstrated that thematic roles share characteristics with nominal concepts in that role/filler featural similarity predicted role/filler typicality, analogous to studies of object concepts (Rosch & Mervis, 1975). This was the first demonstration of this sort with event concepts. Finally, Experiment 3 demonstrated an on-line influence of role features, also a novel empirical result. Resolution of a main clause/reduced relative clause ambiguity was modulated by the presence of adjectives that biased the sentence-initial nominal concept towards being a good agent or patient. This experiment adds to the expanding list of early effects of thematic information on syntactic ambiguity resolution. By so doing, it bears on the issue of how linguistic and extralinguistic knowledge sources are combined and coordinated during on-line comprehension, and provides an additional empirical result that appears to be better suited to the constraint-based model of language processing than to the garden path model.

A major contribution of this paper is that it provides a principled way to capture verb-specific aspects of thematic roles that have apparently been previously considered as linguistically insignificant idiosyncrasies of individual verbs. It does so by claiming that world knowledge of events and their common participants is organised so that it is linguistically relevant, rather than being outside of the knowledge that is directly relevant to language processing. This view is in direct opposition to Chomskyan syntactic modularists (Frazier, 1987; Schlesinger, 1995), but is in accord with the views of cognitive linguists (e.g. Langacker, 1987). In our view, the question of where the knowledge is stored is irrelevant; the relevant questions concern what prompts its computation and how quickly this computation is performed. The claim here is that role knowledge is organised in an efficient manner that promotes its automatic computation whenever a verb is read or heard. Note that the claim is not that every instance of an event is brought to consciousness when the verb is
encountered. In other words, a reader’s consciousness is not flooded with memories of accusing events when she reads ACCUSE. Rather, the claim is that memory is organised in a manner akin to a matrix memory model such as that of Hintzman (1986), or related models such as that of McClelland and Rumelhart (1985), so that one consequence of hearing or reading ACCUSE is the computation of a representation that corresponds to a prototypical accuser. Note that memory models of these types also enable computed representations to be shaded by context. Thus hearing ACCUSE during a movie about a squabbling family might produce somewhat different role representations than during a courtroom drama.

We have described a framework for thinking about how the conceptual aspect of thematic roles might be computed and used in on-line language comprehension. There are, however, a large number of further questions to be addressed. For example, how do role concepts function in conjunction with multiple thematic grids, ambiguous verbs, and obligatory versus optional roles? What is the timecourse of activation and decay of role concepts? Are, for example, patient role concepts more highly activated for transitive- than intransitive-biased verbs? What is the relationship between role concepts and linking rules? We do not attempt to answer these questions here, but instead conceive of them as important questions to investigate in the near future.

Finally, one interesting aspect of thematic processing concerns the distinction between computing thematic fit and contracting features. There are key differences between the features that a filler of a role would be expected to possess (i.e. the features used to compute thematic fit) and those that are contracted to the filler as a result of the event described by the verb. For example, according to the results of Experiment 1, the patient of SERVE should be someone who has characteristics such as ⟨needs help⟩, ⟨is rich⟩, ⟨is powerful⟩, ⟨is important⟩ and ⟨is hungry⟩. The act of serving that person, however, may change a number of those characteristics; that is, some features like ⟨is satisfied⟩ or ⟨is full⟩ might be contracted out (see Schlesinger, 1995, for a detailed description of contracting features). The experiments described above were designed to tap the features that a verb expects a filler to possess, and thematic fit is assumed to be computed over that type of feature. An interesting issue concerns the processes and representations that control the contracting of new features to someone who has been SERVED or CURED.

In conclusion, the present work stands as an example of the potential advantages of combining theoretical notions and empirical techniques from areas of research that are typically treated as distinct. In particular, importing ideas from research in concepts and categorisation into sentence processing provided us with a way to shed new light on an old issue, namely the nature of thematic roles.
REFERENCES


**APPENDIX**

The 20 items used in Experiment 3. Feature pairs are separated by a backslash, with agent-biasing presented before patient-biasing.

1. The shrewd heartless / young naive gambler (who was) manipulated by the dealer had bid a lot more money than he could afford to lose.
2. The honest efficient / dishonest lazy worker (who was) replaced by a computer was forced to find another job.
3. The rude outspoken / thoughtful sensitive student (who was) offended by the comment walked out of the room.
4. The big strong / frail little kid (who was) protected by his brother was never bothered by the bully at school.
5. The persistent intimidating / polite timid beggar (who was) harassed by the police moved to a new location.
6. The respected veteran / young inexperienced employee (who was) trained by the specialist quickly learned how to operate the new equipment.
7. The mean and devious / harmless and gullible college student (who was) tricked by his buddies was upset by their practical joke.
8. The angry aggressive / preoccupied unsuspecting teenager (who was) attacked by the mugger had foolishly tried to cut through the park.
9. The calm quiet / violent unpredictable child (who was) feared by his classmates sometimes fought out on the schoolyard.
10. The large hungry / small defenseless animal (who was) devoured by the lion had been on its own looking for food.
11. The wealthy and generous / penniless and sick old man (who was) supported by his daughter decided that it was time to go to the hospital.
12. The loud screaming / quiet peaceful baby (who was) disturbed by the noise refused to lie back down.
13. The ugly scary / nervous timid man (who was) frightened by the gunshot believed that he was going to die.
14. The daring courageous / helpless injured mother (who was) rescued by the fireman was taken to the hospital for smoke inhalation.
15. The ruthless sadistic / unconscious bleeding spy (who was) tortured by the CIA had thought that he would not be discovered.
16. The young playful / frightened fleeing cat (who was) chased by the dog climbed the neighbour’s tree.
17. The young and aspiring / famous and respected hockey player (who was) idolized by his brother decided that it was time to teach him how to play.
18. The talented artistic / beautiful nude woman (who was) sketched by her friend was happy with how it looked.
19. The dominant authoritative / vulnerable submissive husband (who was) controlled by his wife fooled himself into thinking he was in charge.

20. The obnoxious vulgar / quiet hardworking secretary (who was) annoyed by her boss was constantly complaining and swearing about her work.