Understanding "Rules": When is Behavior Rule-Guided?

Ulrike Hahn (U.Hahn@warwick.ac.uk)
Department of Psychology, University of Warwick
Coventry, CV4 7AL, UK.

Nick Chater (N.Chater@warwick.ac.uk)
Department of Psychology, University of Warwick
Coventry CV4 7AL, U.K.

Abstract

The extent to which human cognition can be understood as rule-based is a classic issue in Cognitive Science and one which continues to provoke heated debate in a wide variety of areas, ranging from Implicit Learning through Inflectional Morphology to the acquisition of reading skills. Despite its centrality, the central notion of "rule" is far from well-defined. This paper examines a central feature of rule-based models, the concept of rule-following, and clarifies its role, its content, and some of the typical fallacies associated with its use.

Introduction

To what extent human cognition is based on rules is a cognitive question of longstanding interest. In the early days of AI, the rule-based nature of human thought was axiomatic; rules no longer have this general, dominant role, but rule-based accounts of particular tasks still abound. Artificial Grammar Learning (Reber, 1989; Brooks & Vokey, 1991; Redington & Chater, 1996) and Inflectional Morphology (Rumelhart & McClelland, 1986; Plunkett & Marchman, 1992; Pinker, 1991; Marcus et al., 1995; Nakisa & Hahn, 1996) are but two areas which are dominated by ongoing debate between proponents of rule-based accounts and supporters of alternative models such as exemplar-based or connectionist accounts. Despite this continued interest in rules, the very notion of rule is one of the most confused within Cognitive Science. This is manifest, to name just one example, in the lack of consensus about whether or not connectionist networks have or embody rules: statements to the effect that they clearly do not, and, hence, offer alternatives to rule-based accounts (Rumelhart & McClelland, 1986; Smith, Langston, & Nisbett, 1992) can be contrasted with the claim that "contrary to rumour, it is not the case that connectionist systems have no rules" (Bates & Elman, 1993, pg. 634).

Conceptual clarification is essential if debate about rules is to have substance. In service of such clarification, this paper focusses on a central aspect of the notion of rule -the dichotomy between behavior which is *guided* by rules as opposed to behavior merely *described* by rules. This distinction is fundamental to what it means for a behavior to be rule-based, yet confusion both about its content and its role prevails.

"Rules" in cognitive contexts: "strong" and "weak" readings

What exactly do researchers mean when they appeal to rules in explaining behavior? We can distinguish two different kinds of usage of "rule", which we will respectively refer to as "weak" and "strong". Examples of weak usage of the term rule are statements such as, for example, a general, behavioral claim that a language learner has succeeded in "mastering the rules of English" or the assumption that infants are born with "rules for looking" which guide their exploration of the visual environment. Statements like these use the term "rule" to refer to an external regularity (of English) or to an internal constraint without making a claim about mental architecture, i.e., without wishing to endorse a particular view about how the external regularity or the innate constraint are internally represented by the agent. Such a weak usage of the term rule in a cognitive context is NOT the focus of the debate about mental rules, nor is it the focus of this paper.

Rather, this debate is concerned with the "strong" use of rule. On the strong reading, speaking of an agent as possessing a rule is a statement about cognitive architecture. It is the claim that an agent has mental representations of a particular representational format, a format which is distinct from other types of mental representation. This stronger, more specific claim lies at the heart of the debates in Artificial Grammar Learning or Inflectional Morphology, where rule-based models are contrasted with exemplar or connectionist accounts.

Rule following

Most importantly, the strong use, on which we will focus below, claims an agent-internal role for the rule. The claim is one about the nature of the representations underlying a particular behavior. Stating that an agent possesses a particular rule is not merely saying that this agent's behavior displays a particular regularity, but rather that this "rule" has a causal role in producing this behavior: the behavior has the regularity it does, *because* the agent posseses the rule in question.

This is commonly phrased in terms of the distinction between rule-guided, or "rule-following", behavior and behavior which is merely conveniently described by rule (see e.g., Marcus et al., 1995). For an example of rule-following, one can think of legal systems and their effect (where documents encoding the law cause particular behaviors such as paying certain amounts of tax), whereas a standard example of rule-

describabeable but not rule-guided behavior is that of the planets—their orbits are well-described by physical laws, but they do not themselves use these laws to guide their behavior.

It is this notion of rule-following, rule-guided behavior, that is implied by the claim that a particular bahavior is rule-based. Stating that performance is based on rules is *not* saying merely that this performance is well-described by a particular rule or rules, but rather that it is an instance of rule-guided behavior.

In the remainder of this paper, we will investigate what this claim really means, how and why it makes sense in the context of cognitive architecture and why it is important. We will also draw attention to some important pitfalls which have the potential to render the empirical debate about rules conceptually vacuous.

The basic intuition distinguishing rule-guided from merely rule-describeable behavior is straightforward enough. Our theory of planetary motion is not a "cognitive theory"; it does not posit mental representations which the planets act on, and correspondingly, we do not assume that planetary motion is "rule-based". Such an account, while logically possible, would be considered ludicrous.

Difficulties start to arise because the same overt behavior can be subject to the rule-guided/rule-describable distinction. Imagine, for instance, drivers stopping at a red light; if I stop at the red light because of my knowledge of traffic regulations I am following the rule; if I stop, totally oblivious to these regulations, because I am intrigued by the sight of a pretty colour, I am behaving in accordance with the rule "if a traffic light is red, then stop", and, hence, my behavior can be *described* by it, but I am not *following* it.

This latter example illustrates why the distinction is relevant in cognitive contexts: only the rule-guided case shows a particular form of a psychological explanation. Only here is an overt behavior linked to a type of explanation, which makes reference to agent internal states, e.g., "I stop, because I know the traffic rule". Merely stating the regularity— "at a red light, the agent stops" lacks this explanatory force.

Regularities and statements of regularities

The preceding discussion implies that in the context of rule-based explanations of cognition "rules" must be sharply distinguished from mere "regularities." A "rule" is not the regularity itself but rather a *statement* of the regularity. In other words, a rule is a representation of the regularity which follows a particular representational format. Different authors disagree on what formats exactly qualify as rules, but all imply rule-following as discussed here.

With the distinction between the regularity per se and the rule as a statement of the regularity firmly in view, we can see that rule-based models offer a kind of explanatory "default account", in the sense that they are conceptually straightforward to generate, a fact which has presumably, greatly contributed to their popularity. Rule-based accounts are straightforward to derive because once a regularity has been observed, we need only posit a representation or statement of that regularity which is agent internal and we have the core of a cognitive account. Given, for example, the observation that the vast majority of English words form a past tense by adding the suffix /ed/—the regularity—one can immediately derive a cognitive account by positing a mental representation of this regularity, roughly "for past tense add /ed/", as an internal rule which speakers are using to generate the appropriate forms (see e.g., Pinker, 1993) and one has a cognitive theory of past tense production.

Of course, just because such an account is conceptually easy to come by, does not mean it is empirically adequate. In all liklihood, the regularity can be exploited in different ways, giving rise to alternative cognitive accounts; in the worst case, the regularity might be spurious—a mere correlate of the "true" underlying cause. Returning to the past tense example, English not only has regular forms, which take +/ed/, but irregulars such as sing/sing, hit/hit, or sleep/slept. These are not just isolated exceptions but often come in families such as sing/ring and they can form the basis for generalization of non-words such as "spling". Hence, a certain regularity holds between phonology of the singular and the type of past tense form a novel word receives. One can exploit this regularity by extracting the relevant phonological features that determine a particular past tense type, collating them into an explicit statement and then using this statement of the regularities to identify the past tense of a new word; on this account novel forms would be generated on the basis of a collection of rules: exceptional rules stating phonological regularities for irregulars and a general rule "add +/ed/" which is used where the exceptional rules fail to apply. Alternatively, a simple "nearest neighbour" strategy (a simple exemplar account) might work just as well: here, a novel word is *always* inflected in the same way as the known word to which it is phonologically most similar. There are no rules at all, neither general nor exceptional. This strategy might work equally well because phonologically close items will naturally share the regularity in question. Nearest neighbour models are "structure mirrors", which reflect the structure present—here, in the English lexicon—but do not *extract* and *explicitly* represent this structures as rules do. They can succeed, because it is ultimately the same regularity that is being exploited, albeit in a different way.

It is this possibility of alternative means of exploiting the same regularity that makes rule-based reasoning an issue and that makes cognitive psychological explanation non-trivial. If rule-based accounts were the only models conceivable, the most challenging task would be that of finding the pertinent regularities. Psychological explanation in terms of mental representations and processes would be reduced to perpetual positing of internal statements of these regularities ("rules") and would add nothing much. Because we know that a plethora of conceptual alternatives exists, however, the discovery of a salient regularity can only ever be the first step for psychological theorizing. This holds even in domains such as natural language syntax where the discovery of the dominant regularities is by no means an easy task. Any description of such a regularity must always be supplemented with and account of how this regularity is exploited. Because rules are only one possible representational format which can deliver

such an account, more must be given than the possibility of a rule-based account; it must be shown that the posited rules are causally efficacious.

Rule following and the causality of representation

Causal efficacy is typically cited as the hallmark of rule-guided behaviour. For instance, Searle, in a critique of Chomsky's (1980) Rules and Representations, holds that, in contrast to rules as used in the natural sciences which merely describe and explain, the use of rules in explanation of human behavior requires that the content of the rule must function causally in the production of the very behavior the rule seeks to explain (Searle, 1980).

First, from what we have said above, it is clear that Searle's position must be disagreed with in one respect: it is not "the content of the rule" which must function causally, but rather the statement (representation) of the rule. Its content, as we have seen is just the regularity in question; but, as we have also seen, this regularity can be exploited in different ways. It is only when it is used by a particular type of representational format that we speak of rule-based accounts; it is this particular type of representation that must function causally, not the regularity.

Second, we must ask what it actually means to "function causally" and how this can be ascertained. Loosely following Chomsky (1986) we assume that we are entitled to hold that an agent is following a rule R if our "best theory" of what the agent is doing, i.e., the best we can construct with the available evidence—invokes a mental representation of R. But this requires further clarification both of what it is a best theory of and what evidence must be taken into account.

We have already seen that rule-guided behavior is about a particular type of explanation. The sort of explanation which such a "best theory" seeks to provide is an account of behavior in terms of mental representations and procedures; For "causal efficacy" we require no more than that the rules are invoked in an explanatory account which involves procedures drawing on representations of these rules and that this explanatory account constitutes our "best theory" available.

Such explanatory accounts in terms of representations and procedures are exactly what researchers engaged in classic rule debates such as Artificial Grammar Learning or Inflectional Morphology are seeking. Most importantly, there is no restriction on what evidence is permissable or relevant.

In our past tense example, evidence for what constitutes the best theory is by no means confined to the ability of the models to produce the right past tense forms. Rather, both models exhibit a whole range of characteristics which give rise to further predictions. For example, they require different learning strategies (rule induction vs. instance storage) and as a result may produce quite different learning profiles: the time course of learning can differ, as might be what is easy and hard. One might be more tolerant of `noise' in the data and so on. *Any* such attributes can be called upon in assessing which theory best fits the data, as well as the desire for parsimony and coherence with other bodies of theory which we bring to the task.

The importance of levels

Additionally, the importance of levels of description must be emphasized. Levels of description are inherent in the context of biology—our theories can invoke brain regions, neurons, or neurotransmitters—as well as in computation where we can reiterate the question of how an algorithm is implemented proceeding downwards from "C-code" to assembly language, to logic gates, to silicon and so on. Hence, the issue of levels is unavoidable for cognitive theorizing. It is pertinent in this context, because accounts in terms of representations and procedures, putative "best theories", might be available at multiple levels as both biological and computational examples suggest.

Specifically, production-rule systems are Turing equivalent, that is, any effectively solvable algorithmic problem can be solved by a production system (Post, 1943). This means, any computation can be made "rule-based" and, as a result, any cognitive theory could be perceived as rule-based if there are no constraints on level. In particular, the nearest neighbour account of the past tense could be implemented using production rules, giving rise to the spurious claim that performance was "rule-based" after all.

Similarly, the constraints on what consitutes "connectionism" seem weak enough to allow implementation of virtually anything and connectionist implementations of "higher-level" cognitive accounts are regularly presented, e.g., Kruschke's (1992) ALCOVE, which implements an exemplar model popular in the categorization literature) or Touretzky and Hinton's (1988) implementation of a production-rule system.

This means commitment to a particular level of description is required. In particular, sweeping contrasts between connectionist and rule-based accounts of cognition, lacking commitment to a particular explanatory level, lack focus and, hence, substance.

The scope of the distinction

Inocuous as our rendition of what it means for behaviur to be rule-guided might look, it has a number of highly desirable properties. First, it applies equally to agent external, "public" rules and to agent internal, "private" rules", i.e., to rules I am told as well as rules I posit to myself;[1] furthermore, it can apply equally to rules, which are formulated in natural language and to "tacit rules" to which we typically have no conscious access. This is because it is defined, generally, in terms of causal efficacy of a representation with requisite format. Again, it does not obviate the need for decision on which formats qualify, but this is a question which itself arises equally for the natural language case and for putative cognitively impenetrable representations.

Second, our rendition of "rule-guided" allows one to see that what is generally treated as one of the many problematic issues about rules is ultimately a general issue of cognitive theorizing. "Rules" are not special: the rule-guided vs. rule-describeable distinction is all about the inference from salient regularities to cognitive models which exploit these regularities. For behaviour to be rule-based, more must be shown than the regularity itself, this "more" being the "causal effi-

[1] This contrasts, for instance, with Quine (1972).

cacy" which we have reconstructed as "explanatory role in our best theory". This "more", however, is a requirement which *any* cognitive account, rule-based or other, must meet. The issue, really, is about certain types of cognitive explanation, not one particular to rules.[2]

What is particular to rules, is the ease with which regularity and rule are confounded and the ready availability of a rule-based cognitive story. Anything can be described by a rule in the sense that any regularity can be stated in a (sufficiently rich) language, in a format which corresponds to our natural intuition of rule. This is simply a fact about language and description. But, given such a description, we can also *always* use this as the heart of a rule-based cognitive model which claims that it is exactly this description (i.e., the statement of a regularity) which is being used by the agent to produce the behavior in question. This is what we referred to as the *default availability* of rule-based models above. Again, however, just because such a model is easy to provide is not sufficient reason for preferring it over competing accounts.

Applying the distinction

We will conclude by discussing two prominent examples which enable us to put the outlined distinctions to use.

The first is a quote from Elman et al.'s, otherwise highly commendable, 1996 book:

"To say that a network does not have rules is factually incorrect, since networks are function approximators and functions are nothing if not rules. So arguments about whether or not networks have rules really do not make much sense. Others have tried to distinguish behaviour which is characterized by rules, and behaviour which is governed by rules. Presumably, in the first case, behaviour only accidentally conforms to a rule, whereas in the latter case the rule has causal effect. Clearly, the behavior of a network is causally connected to its topology and connection weights, so ultimately this is also not an interesting distinction." (pg. 102)

As should be apparent from the preceding sections, the argument based on function approximation is unsound. Even if the function in question is undoubtedly one which we would qualify as a rule[3], all this is saying is that the behavior of the network exhibits a particular regularity, namely that summarized by the function. The argument does not answer the question it is required to, namely whether the network is using a *representation* of that function to produce this behavior. The approximated function is a regularity, the question of rules is about the *means* by which this behaviour is achieved. Equating "having rules" with the behavior of the network (the approximated function) means that planets too "have" the laws of gravitation. But then "having rules" or not ceases to be a question.

Rule-*following* has to be at stake if debate about rules is to have any substance. To show this in a network, requires identifying a representation format which one is willing to call "rule" present *within* the network. It must be decided what kinds of representational formats count as "rules" and whether or not a network exhibits them. This is a task which requires rather more space than we have available here (but see Hahn, 1996; Hahn & Chater, 1998) so we will limit ourselves to a few general comments. The first is that, regardless of what decisions one makes on requisite formats, there is not likely to be a generic answer for all connectionist nets, due to the generality of the criteria defining connectionism and the flexibility they allow. Accordingly, we will limit our own comments to one particular architecture, namely, standard backpropagation networks. Even without attempting any definition of "rule", we can ask what, in such networks, is available as a candidate for "rule". Because rule-following is a matter of causal efficay of a particular representational format, a prerequisite is that putative candidates be representational.

Unfortunately, those parts of the network which undoubtedly are representational in nature fail to be appropriate candidates for other reasons. Both input and output units in a network clearly satisfy the "representational" constraint, but they are not the components that matter: *any* cognitive model assumes representations of inputs and outputs, the debate is entirely about *what is in between*.

The two candidates "in between" are hidden units and weight vectors. Hidden units seem a poor candidate for rules, partly because they simply rerepresent the input in a way that allows problem to be solved by linear mapping from hidden to output. Intuitively, hidden units appear to merely be encoding the results of intermediate calculations involved in mapping from the input to the output. But the question of whether a system has rules seems to be concerned with the *nature of the transformations* between input, intermediate representations and output, not with intermediate representations themselves. Thus, hidden units would seem to be the wrong kind of thing to be candidate rules. Finally, if hidden units representations are candidate rules, this would mean that networks with a single layer of weights could not follow rules. This has the puzzling consequence that a multilayer network follows rules, but consists of a concatenation of single layer networks which do not.

The standard suggestion concerning rules in a network, is that they are encoded not in the hidden units, but in the weight vectors. The question here, however, is whether weight vectors should really be viewed as representational at all. It is common to speak of "knowledge" implicit in the weight vector, but is there reason to assume that a purely causal, non-representational story about weight vectors is not enough? Weights ensure that activation flow is appropriate, i.e., such, that the network gets the mapping right. Why should this be taken to involve a *statement* of the regularity? What additional generalizations about network behavior be-

[2] This has consequences for Kripke's (1982) claim that statements about rule-following are not statements of fact, see Hahn (1996).

[3] It seems questionable whether one would want to call all functions "rules"; functions which succinctly capture a regularity, e.g., x=2y, are perceived as typical rules, an un-computable function which consists of a random mapping between elements of the domain and the range is a function (as long as the mapping is unique) but it has no succinct description. The infinite "look-up" table it constitutes does not accord with our intuitions about "rule". These two examples are merely the extremes of a continuum.

come available if one were to adopt this view? To our knowledge, none have been put forward. This is in strong distinction from a classical rule-following system, like an expert system, where the rules which the system uses in inference provide a completely different level of explanation from the causal story about the workings of the underlying machinery.

This leads to the general question of *why* the rule-guided/rule-describale distinction really matters. From the point of cognitive theory, there appears to be a consistent set of generalizations concerning the behavior which classical rule-based system exhibit: e.g., it is possible discretely to add in extra pieces of knowledge to a rule-based system, which will then interact with previously stored rules; the system can learn by being "told" such knowledge, rather than learning from experience; and it is easy to achieve generalization across extremely disparate items. None of these properties apply to standard backpropagation networks, which have a different set of abilities, learning primarily from experience, where information is accrued incrementally, rather than in discrete packets, and most easily generalizing across similar items. Conversely, rule-based systems have problems in learning from experience, and have difficulty learning "quasi-regular" mappings which involve regular and exceptional cases, particularly if such mappings are governed by subtle effects of similarity. Connectionist networks excel in these domains. Overall, then, it is not clear that any of the important theoretical generalizations associated with rule-based systems carry over to standard backpropagation networks; hence, saying that these networks "follow" rules inappropriately suggests that the two kinds kinds of system share properties on which they actually differ.

There is one further interesting assumption in the Elman et al. quotation, namely, the remark that for merely rule-describable behavior, behavior only "accidentally" accords with the rule. Of course, there need be no accident about the fact that behavior corresponds to the rule; planets do not accidently have the orbits posited by the laws of physics. It is just that they do not use a statement of these laws to compute their orbits.

Assuming that rule-description is always only "accidentally" connected to observable behavior marginalizes the explanatory import that rule-description too can have. Issues of explanatory relevance, seem, to us, to underly the misunderstandings surrounding Chomskian linguistics, our second and final example.

We have repeatedly stressed that rule-following is about a particular kind of explanation and, above, we introduced Searle's comments that the use of rules in psychological explanation is distinct from that in the natural sciences. These issues deserve further elaboration. It is true that, as the case of the motion of planets shows, concise statements of regularities are central to the natural sciences and there are unquestionably perceived as "explanatory". Crucial to the explanatory power is the reduction of a complex behavior to a limited number of variables. Contrary to what Searle seems to suggest, this type of explanation has a role in psychology and Cognitive Science as well. Shepard's Universal Law of Generalization (1987) claims a universal function underlying generalization in humans and a range of non-human species on a variety of tasks. Similarly, "rational analysis" (Anderson, 1990) provides a form of explanation not immediately concerned with mechanism. Most frequently, however, descriptive statements of regularities, "weak" uses of rule, provide a form of explanation which is only partial and, hence, incomplete.

We can illustrate this latter category with a linguistic example, that of the German gender system. Linguistic study and connectionist modelling (see Koepcke, 1993) have isolated phonology as the key factor determining the assignment of gender to German nouns. This has made it clear that German gender, which was previously thought to be arbitrary, is, in fact, highly systematic. A highly complex system and corresponding linguistic behavior are reduced to a single critical variable. Discovering and stating this regularity does "explain" German gender.

From a cognitive perspective, however, we have said that this is always only a first step. Stating the regularity is not a full cognitive account, i.e., an account which explains behavior in terms of mental representations and procedures, simply, because—as seen above—this regularity might be exploited by the cognitive system in a myriad of ways.

The cognitive architecture underlying knowledge of gender might be simple exemplar storage, schemas which abstract families of similar words into more abstract internal representations (Koepcke, 1993) or sets of rules. All of these are conceptually distinct and give rise to different secondary predictions. It is precisely because of these different further predictions that these issues matter to the study of behavior. Finally, this step to internal representations and procedures matters, because it provides a lithmus test for the regularity in question. The "wrong" regularity, e.g., a spurious correlation, will ultimately yield only unsatisfying cognitive accounts, hence, theories in terms of representation and process feed back in to the evaluation of particular descriptive accounts.

All of these issues play a role in the continued debate about Chomsky. Chomskian linguistics, which Chomsky explicitly holds to be concerned with the psychology of the individual (Chomsky, 1980, 1986), aims to answer questions about the nature of linguistic knowledge through the specification of a grammar, i.e., a descriptive account (Chosmky, 1986). Such a grammar is viewed as a putative "best theory" from which we are allowed to infer the entities postulated are "real". This step from description of regularity (grammar) to mental representations is just the step from regularity to rule-based account, which, as we have seen, is an inference which requires further evidence to be justified. While Chomsky does not set out any bounds on "allowed evidence" for what constitutes our best theory, he also shows no positive sign of interest in the type of additional data one needs to resolve architectural in other areas of cognition.

In fact, Chomsky's "Knowledge of Language" (1986) shows considearble disdain for the rule-guided/rule-describeable distinction, a stance which seems to stem from the assumption that all the hard work, at least when it comes to syntax, is discovering and describing the salient regularities. However, if successful, such an account would have

explanatory value, but it would still not provide all the explanation the cognitive psychologits desires. What remains, the second step from describing regularity to best model, is substantive.

Psychology's experience with modelling, the capacity of many models to produce the same overt behaviour, has shown us that the step from regularity to best model is far from trivial; it is not a minor leap from regularity to internal rule. Understanding cognition in terms of mechanisms is both harder and more interesting than once assumed and remains a central issue of practical consequence—for making sense of behavioral data as well as designing artificial intelligent systems. For this reason, the rule-guided/rule-describeable distinction continues to matter.

Summary

We have clarified and justified the distinction between rule-guided and rule-describeable behavior, a distinction which, though classical, continues to prompt misunderstanding. Specifically, we have argued that rule-following implies that our "best theory" of the behavior in question invokes mental representations of the salient regularity governing this behavior represented in requisite format. We have discussed what types of evidence are relevant in determing our "best theory" and stressed the need for commitment to a particular level of description if debate is to remain substantive. We have noted that this criterion applies equally to "public" and "private", explicit and tacit rules. We have also shown how this classic issue surrounding rules is, in fact, not particular to rules; rather it is a general corollary of cognitive theories. It is just that it is particularly confusing in the contexts of rules, because the distinction between regularity and statement of regularity is easily overlooked --a fact which is exacerbated by the fact that we can so easily describe regularities and, from there, generate complete rule-based accounts. We have shown how the rule-guided/rule-describeable distinction relates to explanation in psychology. Finally, we have discussed rule-following in the context of connectionist networks, arguing that standard backpropagation networks seem to lack suitable representational candidates and we emphasized that it is the carry-over of empirical generalizations, gathered in years of computational and experimental research, that is at stake and which makes the distinction between rule-guided and rule-describeable worth preserving.

References

Bates, E.A. and Elman, J.L. (1993) Connectionism and the Study of Change. In M.H. Johnson (ed.) *Brain Development and Cognition*, Oxford: Blackwell.

Brooks, L. and Vokey, J. (1991) Abstract Analogies and Abstracted Grammars: Comments on Reber (1989) and Mathews et al. (1989), *Journal of Experimental Psychology: General*, 120, 316-323.

Chomsky, N. (1980). *Rules and Representations*. New York: Columbia Press.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. Westport, CONN: Prager.

Elman, J.L. , Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, and Plunkett, K. (1996) *Rethinking Innateness: A Connectionist Perspective on Development.* Cambridge, MASS: MIT Press.

Koepcke, K. (1993) *Schemata bei der Pluralbildung im Deutshen*. Tuebingen: Narr.

Kripke, S.A. (1982) *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.

Kruschke, J.K. (1992) ALCOVE: An Exemplar-Based Connectionist Model of Category Learning. *Psychological Review*, 99, 22-44.

Hahn, U. (1996) Cases and Rules in Categorization. Doctoral Thesis. University of Oxford, UK.

Hahn, U. and Chater, N. (1998) Similarity and Rules: Distinct? Exhaustive? Empirically Distinguishable? *Cognition*, 65, 197-230.

Marcus, G.F., Brinkmann, U., Clahsen, H., Wiese, R.,Woest, A., Pinker, S. (1995) German Inflection: The Exception that Proves the Rule. *Cognitive Psychology*, 29, 89-256.

Nakisa, R.C. and Hahn, U. (1996) Where Defaults Don't Help: the Case of the German Plural System.*Proceedings of the 18th Annual Meeting of the Cognitive Science Society*, Mahwah, NJ: Erlbaum.

Pinker, S. (1991) Rules of Language. *Science*, 253, 530-53.

Plunkett, K. and V. Marchman, V. (1991) U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition*, 38, 43-102.

Post, E. (1943) Formal Reductions of the General Combinatorial Decision Problem. American Journal of Mathematics, 65, 197-215.

Quine, W. V. (1972 Methodological Reflections on Current Linguistic Theory. In Harman, G. and Davidson, D. (eds.) *Semantics of Natural Language*, New York: Humanities Press.

Reber, A.S. (1989) Implicit Learning and Tacit Knowledge. *Journal of Experimental Psychology: General*, 118, 219--235.

Redington, M. and Chater, N. (1996) Transfer in Artificial Grammar Learning: A Re-evaluation. *Journal of Experimental Psychology: General*, 125, 123-138.

Rumelhart, D.E. and McClelland, J.L (1986) On learning past tenses of English verbs, In Rumelhart, D.E. and McClelland, J.L (eds.) *Parallel Distributed Processing, Vol 2: Psychological and Biological Models.* Cambridge, MA: MIT press

Searle J.R. (1980) Rules and Causation. *Behavioral and Brain Sciences*, 3, 37-38.

Shepard, R.N. (1987) Toward a universal law of generalization for the psychological sciences. Science, 237, 1317-1323.

Smith, E.E., Langston, C., Nisbett, R.E. (1992) The Case for Rules in Reasoning. *Cognitive Science*, 16, 1-40.

Touretzky, D.D. and Hinton, G.E. (1988) A distributed connectionist production system. *Cognitive Science*, 12, 423-466.