

## How Do We Tell an Association From a Rule? Comment on Sloman (1996)

Gerd Gigerenzer

Max Planck Institute for Psychological Research

Terry Regier

University of Chicago

S. A. Sloman's (1996) intriguing argument for separate associative and rule-based reasoning systems is unfortunately damaged by a certain amount of slack in the distinction he makes between these two posited mental mechanisms. The authors suggest that the distinction could be sharpened by overt reference to explicit models of associative and rule-based processing. They also point out that "simultaneous contradictory belief," which Sloman takes as evidence for separate associative and rule-based systems, need not be interpreted in this fashion. It may also signal a number of other things, including the presence of linguistic ambiguity (as in the Linda problem), competing lines of formal reasoning (as in the Wason selection task), and unclarified assumptions (as in the 3 doors problem).

In 1973, Allen Newell, torn between being distressed and content with the state of research on information processing, entitled a commentary "You Can't Play 20 Questions With Nature and Win." Newell's distressed half fear that when behavior is explained in terms of binary oppositions—serial versus parallel, grammars versus associations, nature versus nurture, and so on—"clarity is never achieved" and "matters simply become muddier and muddier as we go down through time" (pp. 288–289). His concern was that despite brilliant experiments, there would be little cumulation in theory.

In a stimulating article, Sloman (1996) tackles one such opposition head-on. In particular, he argues that the mind contains two separate reasoning systems: one that operates associatively and another that operates in a rule-based fashion. In support of this hypothesis, he reinterprets a broad and suggestive collection of results that appear to lend some credence to his thesis. We appreciate the potential usefulness of analyses of this sort and of dual-process ideas generally, particularly as first steps toward eventual models. In fact, we suspect that a somewhat clearer formulation of the proposed distinction may prove to be helpful in understanding human reasoning. However, we are left uneasy by the inchoate nature of the particular distinction drawn. By proposing a conceptually blurred dichotomy, Sloman seems to buttress his own thesis, but he also underscores Newell's (1973) concern that explanation in terms of binary oppositions can be a risky business. We try to substantiate this criticism and then consider how the distinction might be tightened a bit.

Sloman (1996) views the *associative system* as primarily reflecting "similarity and contiguity" (p. 4); it operates "reflexively"; and it "often deals in concrete images" but "can also deal in abstract concepts" (p. 16). Associative processing is generally unconscious—but then rule-based

processing can be, too. The rule-based system is characterized by “productivity,” “systematicity,” “variables,” and “sequential” operation, and it captures the “logical, hierarchical, and causal-mechanical” (p. 6) structure of the environment, as opposed to its “statistical” structure. The associative system produces “quick and dirty” answers, based on heuristics such as representativeness or availability, whereas the rule-based system can provide explanations of the environment. Sloman points out that this distinction is similar to Smolensky’s (1988) distinction between an intuitive processor and a conscious rule interpreter, to Hinton’s (1990) distinction between intuitive and rational processing, to Schneider and Shiffrin’s (1977) distinction between automatic and controlled processing, to Evans’ (1989) distinction between a perceptually-based matching process and a linguistic-logical process, and to Freud’s (1913) distinction between primary and secondary processes. Sloman could also have added Shastri and Ajjanagadde’s (1993) distinction between reflexive and reflective reasoning, as well as a number of others.

The problem is that these distinctions are not all the same. Some of them provide precise and ultimately testable characterizations, but this clarity is lost when they are superimposed one on top of the other in an all-embracing associative versus rule-based distinction. For instance, in discussing Smolensky’s (1988) intuitive processor, Sloman seems comfortable with the assignment of numerous soft constraints to the associative system and hard constraints to the symbolic rule-based system (p. 8). This however causes trouble. Consider, for example, probabilistic reasoning systems such as Bayesian belief networks (Pearl, 1988). These systems use many soft constraints that “cooperatively contribute to a state that is maximally consistent with all the units and associations” (Sloman, 1996, p. 8) and should therefore be expected to reside on the associative side of Sloman’s cognitive divide. However, they are also rule based, by other criteria: The constraints hold among symbolically expressed propositions, and these mechanisms explicitly follow (rather than merely conform to) the rules of probability. Thus, the existence of such systems forces a choice between a distinction cast in terms of hard and soft constraints and one cast in terms of rule following.

More generally, Sloman’s (1996) treatment of his distinction in terms of probability and statistics is problematic. He claims that the associative system makes “predictions that approximate those of a sophisticated statistician” (p. 4). This statement grants the associative system an impressive breadth of coverage and is inconsistent with the more cautious assertion that this system is sensitive merely to similarity and temporal contiguity. Whatever similarity and temporal contiguity are, they are certainly not coextensive with statistics. The fluid nature of the distinction is also highlighted by the fact that the representativeness heuristic, firmly placed in the associative system, can be easily expressed as a likelihood in the formal (rule-based) language of probability (Gigerenzer & Murray, 1987, pp. 153–157). Again, this forces a choice: Either one considers representativeness to be associative because it is similarity based, even if it is computed by explicitly following a formal rule, or one concedes that it may be rule based if such a rule is explicitly followed.

Sloman sees another incarnation of his distinction in the work of Hinton (1990), who differentiated between inferences that take place in a single settling of a connectionist network and those that require a sequence of such settlings with different variable bindings each time around. Sloman classifies the former as associative and the latter as rule based. Hinton’s distinction is one you can sink your teeth into, but it is not the same as the other distinctions to which Sloman ties it. For example, in Tversky and Kahneman’s (1983) Linda problem, the answer which Sloman classifies as rule based relies on the single application of only one rule: The probability of a conjunction cannot be greater than that of either of its conjuncts. There is no seriality involved. This is at odds with the notion that the rule-based system is necessarily sequential in nature. The prob-

lem then is that Sloman's distinction is so broad that it forsakes the precision of the constituent subdistinctions that it seeks to encompass and glosses over their incompatibilities.

To be fair to Sloman, it should be added that ambiguity in modeling seems to be a part of our field. Consider, for example, the representativeness heuristic, invoked to explain the conjunction fallacy (Tversky & Kahneman, 1983), as one of the prime examples of Sloman's (1996) associative system. It is understandable that when this heuristic was first introduced around 1970, it was loosely defined. Yet, more than 20 years and many experiments later, representativeness is still often treated as a vague and undefined notion (Shanteau, 1989; Wallsten, 1983). Whereas the work of Smith and Osherson (1989) represents a refreshing deviation from this trend, in other cases representativeness seems to be hardly more than a redescription of the cognitive illusion it purports to explain (Gigerenzer & Murray, 1987, chapter 5). This, together with other underdefined explanatory concepts such as *availability*, *anchoring*, *familiarity*, and *salience*, can be used to account for an impressively wide array of behaviors post hoc.

To continue being fair to Sloman, it should also be acknowledged that he clearly recognizes that the distinction between the two systems is a slippery one: "Any apparently associative process can be described as a rule based because of the representational power of rules." (Sloman, 1996, p. 11) Despite this slipperiness, however, there is the potential for a more cleanly specified distinction, and it is contained in Sloman's presentation. He is aware of the fact that the term *associative* in and of itself is not particularly enlightening, so he gives an initial working definition: The associative system is sensitive to similarity and temporal contiguity. This is a promising beginning, but it gets muddled when the associative system is also viewed in terms of Smolensky's (1988) and Hinton's (1990) connectionist proposals and in terms of statistics as a whole. As we have seen, these definitions are not always mutually compatible. Sloman's best bet may be to make the "similarity plus temporal contiguity" definition more precise and then stick to it. One possible way to do this would be to adopt a specific model of similarity, such as Tversky's (1977) or Smith and Osherson's (1989), in conjunction with a similarly precise model sensitive to temporal contiguity (Hebb, 1949). In fact, something one level more abstract may suffice—such as the notion of similarity as a function of featural overlap. This formalization of the concept of similarity would not necessarily correspond to similarity judgments as empirically elicited (Bassok & Medin, 1995), but at least the feature-overlap formalization is clear. One would know when an inference is associative and when it is not: It is associative if and only if it can be accounted for by featural overlap or temporal contiguity, regardless of the hardness or softness of the constraints involved and regardless of the seriality or nonseriality of the reasoning process. Indeed, sometimes Sloman (1996) does seem to have such a conception in mind. In discussing the dependence of similarity judgments on apparently nonfeatural influences such as hierarchical and causal structure (p. 16) and the implications this has for his distinction, he suggests a means to retain a featural view of similarity despite such evidence. In addition, he does take the trouble to point out the difference between his distinction and some related ones (Sloman, 1996, p. 19). But his discussion of the connectionist and statistical aspects of the associative system unfortunately cloud the issue: We are left with several imperfectly overlapping definitions, rather than just one clear one.

Although Sloman concedes that the distinction between associative and rule-based systems is a troubled one, he feels that he can find empirical evidence for it, in the form of "simultaneous contradictory belief" (1996, p. 11; Criterion S). Let us consider this argument.

## Empirical Evidence for Two Systems of Reasoning

Sloman's (1996) Criterion S is met when a person simultaneously holds two contradictory beliefs. The idea is that this occurs when the associative and rule-based systems arrive at conflicting responses to a situation. We argue that whereas Criterion S may sometimes provide evidence for such a collision, it may also indicate a number of other states of affairs. For the sake of this argument, we assume that the associative system is defined, as suggested earlier, in terms of featural overlap and temporal contiguity alone.

One of Sloman's flagship examples of Criterion S is the confusion some participants experience when considering the well-known Linda problem (Tversky & Kahneman, 1983):

Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which of these two alternatives is more probable?

Linda is a bank teller. (T)

Linda is a bank teller and is active in the feminist movement. (T&F) (p. 297)

A majority of participants—some 80% to 90%—choose T&F as more probable. Tversky and Kahneman (1983) argued that this is an error in reasoning: Because the mathematical probability of a conjunction cannot be greater than that of one of its conjuncts, the correct answer is T. Their explanation of participants' behavior is that people seem not to reason by the laws of probability and that they use similarity instead, a strategy termed the *representativeness heuristic*. This similarity-based heuristic is considered by Sloman (1996) to be an element of the associative system, whereas explicit following of the conjunction rule is considered to be rule based.

For Sloman (1996), the critical point here is that the same participants who seem to reason on the basis of similarity also acknowledge that the other answer is correct; yet their original response continues to seem valid to them. They simultaneously believe the associative and the rule-based answers, having arrived at the first themselves, and having been persuaded of the validity of the second by the experimenters. We use the term *persuaded* deliberately—as we shall see, there is in fact no single correct answer to the Linda problem.

The Linda problem seems at first blush to provide support for Sloman's (1996) contention that there are two reasoning systems, one associative and the other rule based, and that the two can be perceived clearly when they are in conflict. However, we argue that this is not necessarily so, that the associative response in this case can be accounted for at least in part by the linguistic ambiguity of the instructions that the participants read, rather than by similarity. We appreciate that some researchers have looked for and did not find an effect of linguistic ambiguity in the Linda task (Crandall & Greenfield, 1986), but other results in the literature provide persuasive evidence that there is in fact such an effect. In particular, a variety of semantic and pragmatic inferences are drawn from the instructions in the Linda problem, and these inferences radically affect the interpretation of the instructions. Whereas some 10–20% of the participants seem to read the task as a straightforward problem in elementary probability, many other participants draw one of several conversational implicatures (Adler, 1991; Dulany & Hilton, 1991). For instance, some 30–40% seem to draw the “T  $\rightarrow$  T & not F” implicature, that is, “Linda is a bank teller” implies that she is a bank teller and not active in the feminist movement (Hertwig, 1995; Tversky & Kahneman, 1983); some 18% seem to draw the “T&F  $\rightarrow$  F given T” implicature (Hertwig & Chase, 1995). Both of these implicatures give firm foundations for the T&F response, meaning that this response need not be attributed to a similarity-based associative pro-

cess. This, in turn, means that Criterion S need not be indicative of a clash between separate associative and rule-based systems.

More generally, the presence of two conflicting responses need not imply two such systems of reasoning. Consider the Wason selection task (1966), also used by Sloman (1996) to illustrate Criterion S. Here proper reasoning about a conditional statement is considered to be reducible to propositional logic, and deviating responses are attributed to matching biases, availability, and other elements of what Sloman labels the associative system. Experimenters are sometimes able to persuade participants that their original choices were in error, at which point the participants meet Criterion S: They continue to believe in the validity of their original choice, but they also acknowledge the logic behind the experimenter's argument. This can be taken as evidence for Sloman's two-system idea as long as one is comfortable attributing the participants' original responses to similarity. Recently, however, researchers have realized that propositional logic is not the only way to model proper reasoning in the selection task (including the so-called abstract version) and have pointed out that there exist other formal systems, such as Neyman-Pearson theory (signal detection theory) or Bayesian reasoning (Oaksford & Chater, 1994), which give answers that are inconsistent with propositional logic and which in the past have been attributed to associations or availability. Similarly, Birnbaum (1983) has shown that the apparently irrational participants' modal responses in the cab problem are consistent with a response derived from the Neyman-Pearson theory of hypothesis testing. The point then is that such different and inconsistent responses can be accounted for by a conflict between alternate formal, rule-based systems, rather than between an associative system and a rule-based one.

Finally, consider the well-known three doors problem from "Let's Make a Deal." Monty Hall shows you three doors: A, B, and C. He tells you there is a grand prize behind one of the doors and booby prizes (goats) behind the others, and he asks you to pick one. You pick Door A. He then opens Door C, and you see a goat. Monty now gives you the option of switching to Door B. The question is, should you? The answer depends on what assumptions are made regarding Monty's actions. The literature lists a number of rationales for switching and a number of rationales for not switching, each of which is valid under a different set of assumptions (Falk, 1992; Shimojo & Ichikawa, 1989). Many people experience Criterion S when considering the three doors problem: There are two perfectly reasonable-seeming incompatible solutions vying for acceptance, and there are solid rationales for each. Do we now want to posit two separate reasoning systems? It seems to us that we do not—it is far simpler to just note that each answer is valid under its own set of assumptions. Thus, another possible cause of Criterion S is the lack of specification of critical starting assumptions.

Just what does Criterion S tell us? Although it is true that Sloman's (1996) examples all seem at first to have associative and rule-based halves, the availability of alternate explanations for the phenomenon is somewhat troubling. Criterion S may result from linguistic ambiguity (as in the Linda problem), competing lines of formal reasoning as in the Wason selection task (Wason, 1966), unclarified assumptions (as in the three doors problem), or possibly from two separate reasoning systems. The key problem then is that Criterion S may signal any of a number of things, not just the existence of separate associative and rule-based reasoning systems.

## A Final Word

Let us try to tie things together. Whereas Sloman's (1996) overall case for two reasoning systems is appealing, it could be sharpened in some respects. On the one hand, we feel that the conceptual distinction Sloman draws between associative and rule-based reasoning systems is ambiguous in nature. On the other hand, one of our criticisms of the reasoning behind the Criterion S argument is that in at least some cases, linguistic ambiguity may account for responses that were attributed to the associative system. In fact, there is an intriguing phenomenon here: Whereas we as psychologists tend to assume that participants will understand a single precise meaning from our instructions, we ourselves traffic in explanatory terms such as *association* and *similarity* that are too often not specified with the same semantic precision we expect of our participants. Thus, ambiguity is cropping up where it should not (in theory) and not cropping up where it should (in considerations of participants' interpretations). If this observation can bring about a Criterion-S-like sensation of inconsistency in the mind of the reader, this article will have served its purpose.

## References

- Adler, J. (1991). An optimist's pessimism: Conversation and conjunction. In E. Eells & T. Maruszewski (Eds.), *Studies on L. Jonathan Cohen's philosophy of science* (pp. 251–282). Amsterdam: Rodopi.
- Bassok, M., & Medin, D. L. (1995). Structural and thematic alignments in similarity judgments. *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*, 460–465.
- Birnbaum, M. H. (1983). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology*, 96, 85–94.
- Crandall, C. S., & Greenfield, B. (1986). Understanding the conjunction fallacy: A conjunction of effects? *Social Cognition*, 4, 408–419.
- Dulany, D. E., & Hilton, D. J. (1991). Conversational implicature, conscious representation, and the conjunction fallacy. *Social Cognition*, 9, 85–110.
- Evans, J. St. B. T. (1989). *Bias in human reasoning*. Hillsdale, NJ: Erlbaum.
- Falk, R. (1992). A closer look at the probabilities of the notorious three prisoners. *Cognition*, 43, 197–223.
- Freud, S. (1913). *The interpretation of dreams*. London: George Allen & Unwin.
- Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.
- Hebb, D. (1949). *The organization of behavior*. New York: Wiley.
- Hertwig, R. (1995). *Why Dr. Gould's homunculus doesn't think like Dr. Gould: The "conjunction fallacy" reconsidered* (Doctoral dissertation, Universitaet Konstanz, Germany, 1995). Konstanz, Germany: Hartung-Gorre Verlag.
- Hertwig, R., & Chase, V. M. (1995). *The role of response mode in probabilistic reasoning*. Unpublished manuscript, University of Chicago.
- Hinton, G. E. (1990). Mapping part-whole hierarchies into connectionist networks. *Artificial Intelligence*, 46, 47–76.
- Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283–308). New York: Academic Press.
- Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, 101, 608–631.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Morgan Kaufmann.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1–66.
- Shanteau, J. (1989). Cognitive heuristics and biases in behavioral auditing: Review, comments and observations. *Accounting Organizations and Society*, 14 (1/2), 165–177.
- Shastri, L., & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences*, 16, 417–494.

- Shimojo, S., & Ichikawa, S. (1989). Intuitive reasoning about probability: Theoretical and experimental analyses of the "problem of three prisoners." *Cognition*, 32, 1–32.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3–22.
- Smith, E. E., & Osherson, D. N. (1989). Similarity and decision making. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 60–75). Cambridge, England: Cambridge University Press.
- Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11, 1–23.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327–352.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 293–315.
- Wallsten, T. S. (1983). The theoretical status of judgmental heuristics. In R. W. Scholz (Ed.), *Decision making under uncertainty* (pp. 21–39). Amsterdam: Elsevier.
- Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology* (pp. 131–151). Harmondsworth, England: Penguin.