Bugs and Biases: Diagnosing Misconceptions in the Understanding of Diagrams

James E. Corter (corter@tc.edu) [a]
Yun Jin Rho (yjr2101@columbia.edu)[a]
Doris Zahner (dzahner@stevens.edu) [b]
Jeffrey V. Nickerson (jnickerson@stevens.edu) [b]
Barbara Tversky (tversky@tc.edu) [a]

a Department of Human Development, Teachers College, Columbia University
525 W. 120th St, New York, NY 10027 USA
b Howe School of Technology Management, Stevens Institute of Technology
Castle Point on Hudson, Hoboken, NJ 07030 USA

Abstract

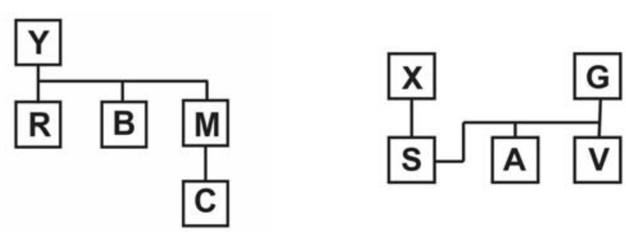
Errors in understanding diagrams come from inappropriate ways of interpreting the diagrams, ways that have a basis in the visual structure of the diagram. In the case of information systems diagrams, these misconceptions can be diagnosed through a Bayesian causal network, in which latent misconceptions are inferred from a simple test on paths in diagrams. The misconceptions are related to surface errors in a merely probabilistic manner. Nonetheless, a model derived from one diagram can be used to make accurate predictions about errors on isomorphic diagrams. The technique can be used to assess misconceptions, including biases and bugs, and may be applied to many different problem domains. There are also pragmatic implications to this work: the domain of this application, that of a local area network, permeates information systems, and the diagnosis and correction of misconceptions will be helpful for those involved in information systems education, design, and trouble-shooting.

Keywords: diagram understanding; Bayesian causal networks; educational diagnostics; design of information systems.

Introduction

People rely on diagrams in all realms of life, maps to navigate, visual instructions to operate a camera, diagrams in education, science, business, design, and more. One reason that diagrams are effective is that they map visuospatial aspects of the world to visuospatial characteristics of paper, readily allowing visuospatial inferences (Tversky, 2001). Although they appear simple, especially to experts, beneath the surface they depend on simplifications, spatial analogies, and social conventions that are usually learned implicitly. Yet this learning does not always happen -- errors occur even for experts and even in the simplest of diagrams, networks of nodes and links (Corter, et al., 2008). Paradoxically, a possible source of error is exactly the visuospatial characteristics of the diagrams that make them effective.

There are many inferences that can be made from even simple diagrams. Consider Problem 1 in Figure 1. A variety of inferences are prompted by the objects in boxes, the lines among them, and their arrangement in space and along the lines. You can infer that the diagram concerns several objects, Y, R, B, M, and C; you may notice that Y is above all the others and C below, that R is closer to B than M, and that B is between R and M. If this were a map, say of buildings on a campus, it could also be used to infer routes; you could infer that the shortest path from Y to C would pass R, B, and M in that order. However, because this is a diagram of components of an information system, valid route inferences differ from those appropriate to maps. One reason they are different is a special convention used in diagrams of information systems. In information systems, there are often clusters of components such as computers that are completely interconnected through Local Area Networks, or LANs. Drawing lines between all possible connections for even a small number of components would make a diagram unreadable. Instead, the interconnected components are depicted as a set of objects dangling from a single line, RBM in Problem 1 and SAV in Problem 2. Thus, a shortest route from Y to C bypasses R and B, passing through only M, called a *bridge* node. This yields the correct path YMC.



Problem 1 Problem 2

Figure 1: Two topologically equivalent networks, each depicting a set of completely connected nodes (YRBM or GVAS) connected by a bridge node (M or S) to another component (C or X).

Generating correct paths from such diagrams requires both declarative and procedural knowledge. The conceptual knowledge needed includes both general knowledge of paths in graphs and in space, and specific domain knowledge about the conventions for representing such systems with diagrams. For example, students must first understand that any path determines a fixed ordering of components along the path; next, they need to determine that correct ordering of components. To determine the correct orders, they need to understand the LAN convention, and to be able to differentiate LAN components from bridge devices that connect subsystems. When these crucial concepts are not mastered, students operate under misconceptions that cause systematic errors.

The procedural knowledge required in the paths task includes strategies for avoiding execution errors. For example, in a task requiring participants to generate all possible shortest paths in a diagram, choice of an enumeration strategy can make a difference in performance. Also, biases in performance make some execution errors more likely than others. Some of these biases arise from the fact that system diagrams representing the topology of a system are typically depicted embedded in the Euclidean plane. When people attempt to extract topological information from a diagram, they cannot avoid being influenced by irrelevant spatial aspects of the diagram, leading to predictable biases (Corter et al., 2008).

This analysis suggests five kinds of misconceptions that might cause errors when participants are asked to generate all possible shortest paths of information flow for Network 1 and Network 2. If participants extract the objects from the diagrams but not that the lines indicate that objects are connected in an order, we say that they show the *no order* misconception. If they interpret lines as indicating order, but do not understand the LAN convention, that all components within a LAN are directly connected (e.g., they propose Y-R-B-M-C as a shortest path), then we say they show the *LAN-as-path* misconception. A companion aspect of interpreting the LAN convention is to understand that there is often a bridge node, an entry-way, into and out of a completely connected component. Participants may fail to notice a bridge node, say, omitting C-M-Y as a shortest path, which indicates the *omit bridge* error. Conversely, a natural visual cue to an entrance is an extreme position, at an edge, as for nodes Y and C in Network 1. If participants misinterpret extreme positions as bridge nodes, for example proposing Y-R-B as a shortest path, then they show the *commit bridge error.* Finally, our previous research (Corter et al., 2008) has shown that participants often read and generate paths in left-right reading order. Thus, they may fail to generate all shortest paths, most frequently by omitting reverse-reading-order paths, demonstrating a *reading order* bias. A relatively rare misconception leading to a reading-order bias is to fail to understand that one must list backward paths as well as forward paths.

To summarize, we have identified five important misconceptions, execution errors, and biases in the paths task:

B1. *no order* misconception: S fails to understand that paths are ordered (e. g., commits CYB).

B2. *LAN-as-path* misconception: S believes that a path through a LAN visits intervening nodes in sequence, like a physical path (e.g., commits YRBMC)

B3. *omit bridge* misconception/error: S fails to recognize or list a true bridge node (e.g., commits BC or omits BMC)

B4. *commit bridge* misconception/error: S infers a nonexistent bridge node, or selectively commits the *LAN-as-path* error (e.g., commits YRB)

B5. *reading-order* misconception/bias: S lists <u>only</u> forward paths (misconception); or omits backwards paths more often (bias)

Knowing which of these errors occur consistently within and across participants is crucial to understanding how people interpret diagrams and consequently how to design and teach them. However, many errors are ambiguous with respect to cause. Omitting CMY is an example; it could occur from ignoring a bridge node or from the LAN misconception or from a tendency to omit backwards paths. Furthermore, the effects of misconceptions on procedural skills (i.e., the appearance of "bugs") are often highly variable, appearing only probabilistically (vanLehn, 1990). For this reason, Bayesian networks have been proposed for the diagnosis of specific bugs in procedural skills (Lee, 2003; Lee & Corter 2009). Bayesian networks (e.g., Pearl, 1988; Jensen, 1996; Korb & Nicholson, 2004) are an ideal tool for combining evidence from probabilistic indicators of latent traits or states, specifically for identifying signal (e.g., specific misconceptions) in noise (e.g., execution errors or "slips"). The Bayesian approach also seems especially promising for diagnosing *biases* in performance, which we define as systematic, stochastically predictable execution errors. More generally, Bayesian networks have been successfully applied to assess specific knowledge components or subskills in educational domains (e.g., Beland & Mislevy, 1996; Conati et al., 2002; Pardos et al., 2007; vanLehn & Martin, 1998; Williamson et al., 2006).

Here, we describe a Bayesian method for diagnosing misconceptions and biases in reasoning about paths in diagrams. Our intent in developing this method is both to identify specific difficulties in individual students, allowing specific correctives, and to identify general difficulties that may be overcome by better design or explicit instruction. Both the methods and the errors have broad implications, because such diagrams are widely used.

METHOD

Materials. A test was designed to measure the ability to reason about paths in diagrams. The two diagrams in Figure 1 were presented, and participants were asked to check which of 19 proposed paths were valid shortest paths Each test question was constructed by selecting 10 correct minimal paths, and 9 potential commission errors actually observed in a previous study using this problem in an open-ended task where participants were instructed to list all shortest paths (Corter et al. 2008). These 19 correct and incorrect paths were used as answer alternatives. Thus, for each depicted network, participants could make errors of omission (by omitting correct paths listed as answer alternatives) or commission (by checking incorrect paths listed as answer alternatives).

Participants. Participants (N=195) were solicited via a posting on a public website asking for participants to: "Solve problems related to diagrams (knowledge of information systems is helpful)." They were compensated with a nominal stipend.

Modeling. We constructed a Bayesian network for each diagram, designed to diagnose a set of specific misconceptions we had identified by analysis of previous data. This initial network incorporates a set of hypotheses about the skills needed for making inferences from diagrams, about potential misconceptions that a participant might have, and about the surface errors that these misconceptions might cause.

For each type of misconception identified above (including execution errors and biases) that a participant might exhibit, the set of observed omission and commission errors that provide positive evidence was coded. Again, the misconceptions are:

B1. *no order* misconception

B2. *LAN-as-path* misconception

B3. *omit bridge* misconception/error

B4. *commit bridge* misconception/error

B5. *reading-order* misconception/bias

The model was trained on data from participants who completed the two test questions concerning shortest paths in two isomorphic diagrams (Figure 1). Participants who checked only a single path in a question were regarded as uncooperative respondents who participated only nominally in order to earn the stipend, or as having misunderstood the question. Thus, data from these participants were eliminated, leaving 128 valid observations. Each participant's data vector for each of the two questions was entered as evidence into a Bayesian network, and the diagnosed misconceptions from the two questions were compared - a form of test-retest reliability. Inconsistent or unreliable diagnosis of several of our hypothesized misconceptions was taken as evidence that the conceptual/cognitive model was wrong. These unreliably diagnosed misconceptions were eliminated from the model, which was re-estimated.

Defining and Training the Networks

For each problem, a 2-layer Bayesian network was defined, with five parent nodes representing the hypothesized bugs, and 19 leaves representing the possible omission or commission errors. The Bayesian networks for the two problems were not identical, even though the two presented diagrams were topologically equivalent, because the definition of "backward" path differs for the two diagrams, so different links were specified from specific omission or commission errors to the node for reading order bias for the two networks. The full Bayesian network defined for Problem 1 is shown in Figure 2.

Algorithmically, our approach relies on standard belief propagation techniques (Pearl 1988). The HUGIN system (Anderson et al., 1989) was used to instantiate the Bayesian network and perform these computations. The network parameters (the simple and conditional probabilities relating misconceptions and errors) were set to reasonable initial values, then expectation-maximization (EM) learning was applied. The trained networks were used to diagnosis the presence of the hypothesized latent misconceptions and biases for each participant.

The observed proportions of participants who made specific omission and commission errors on the two

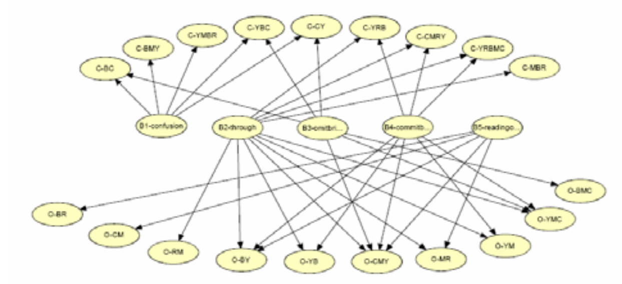


Figure 2: The Bayesian network for Problem 1.

network problems are shown in Table 1. There were more omission errors than commission errors. The most frequent commission errors were YRB, YRBMC and MBR for Problem 1 and GVA, GVASX, and SAV for Problem 2, each committed by approximately 35-40% of participants. CMRY and XSVG were committed by 11% and 17% of participants, respectively.

Training of the Bayesian networks resulted in estimated posterior probabilities for the diagnosed misconceptions (Table 2) that differed radically from the 50% base rates used as priors. The most common misconception was B2 (the LAN-as-path error), while misconception B1 (no-order) was quite rare. For misconceptions B1-B4, similar mean posterior probabilities were obtained using Problems 1 and 2, demonstrating good reliability for estimation of these misconceptions. However, the estimated proportions of participants showing the reading-order bias (B5) did differ between the problems.

Table 1: Proportions of participants making specific omission and commission errors for the two problems.

Problem 1	Omissions:	BR	CM	RM	BY	YB	CMY	MR	YM	RB	MC	YMC	BMC
		0.07	0.05	0.41	0.38	0.36	0.48	0.42	0.41	0.05	0.08	0.53	0.29
	commissions:	BMY	YMBR	YBC	CY	YRB	CMRY	YRBMC	MBR	BC			
		0.03	0.04	0.05	0.06	0.36	0.11	0.38	0.41	0.05			
Problem 2	Omissions:	AV	XS	VS	AG	GA	XSG	SV	GS	VA	SX	GSX	ASX
		0.02	0.09	0.36	0.29	0.29	0.45	0.34	0.34	0.06	0.06	0.45	0.30
	commissions:	ASG	GSAV	GAX	XG	GVA	XSVG	GVASX	SAV	AX			
		0.03	0.05	0.05	0.06	0.38	0.17	0.34	0.37	0.06			

Table 2: Estimated base rates of the misconceptions after EM learning

network	problem	B1	B2	B3	B4	B5
5-bug	P1	0.08	0.35	0.25	0.20	0.13
	P2	0.09	0.28	0.24	0.25	0.08
4-bug	P1	0.04	0.35	0.24		0.13
	P2	0.10	0.30	0.24		0.06
3-bug	P1	0.06	0.36			0.14
	P2	0.06	0.29			0.06

The row of Table 3 labeled “5-bug” shows the test-retest reliabilities of the five latent variables (the posterior probabilities for the five misconceptions), measured as the correlations of the posterior probabilities of the misconceptions between the two problems, P1 and P2. All misconceptions are being diagnosed reliably ($p<.05$), although the correlations for B5 are lower.

Table 3: Correlations of posterior probabilities of the latent misconceptions across the two isomorphic problems.

network	B1	B2	B3	B4	B5
5-bug	.83	.80	.83	.79	.42
4-bug	.37	.75	.84		.35
3-bug	.82	.82			.34

The use of Bayesian networks or any other diagnostic testing method cannot be fairly evaluated if the cognitive model of the domain is incorrect. We had hypothesized the existence of misconceptions and biases B1-B5 based in part on a theoretical analysis of the knowledge and skills required in the domain, but also based on evidence from previous studies in our lab (e.g., Corter et al., 2008). Because of inconsistent evidence for misconceptions B3 and B4 in previous analyses, we compared the 5-bug model with two more parsimonious models, eliminating first B4, then B3 from the Bayesian network for each problem, and retraining the more compact networks on the same data. The rows of Table 3 labeled “4-bug” and “3-bug” shows that the remaining misconceptions have good test-retest reliability in the reduced models too (though less so for B5).

The relative fit of these three models can be compared using model comparison statistics such as AIC or BIC. The 5-bug, 4-bug, and 3-bug models are compared in Table 4. In this table AIC is defined as LL - k, and BIC as LL - k/2 log(n), thus larger values are desirable. By both criteria, the 5-bug model provides the best fit, even adjusting for the additional number of parameters.

Table 4: Model-fit statistics for the three models.

network	problem	AIC	BIC	LL
5-bug	P1	-765.53	-889.60	-678.53
	P2	-700.53	-833.15	-607.53
4-bug	P1	-810.37	-895.93	-750.37
	P2	-747.13	-838.40	-683.13
3-bug	P1	-834.00	-899.60	-788.00
	P2	-811.03	-880.91	-762.03

One way to assess the validity of the misconception diagnoses is to relate them to external criteria. A criterion available here was performance on a third question administered to participants, a problem that also involved path inferences. This problem asked participants to identify the shortest path (as measured by topological distance) between two nodes in a graph. The question was designed so that one alternative answer gave a path between the two nodes that was shortest in terms of Euclidean distance (the most compelling incorrect “distractor” answer) and another was shortest in terms of topological distance (the correct answer). The posterior probabilities for misconception B2 (but not for the other misconceptions) were negatively and significantly correlated with performance on this related problem.

We also investigated multiple diagnoses: in the 5-bug model only 15 participants were diagnosed with more than one type of misconception, and the most common patterns were the combination of B2 with B3 (n=8), followed by B1 with B4 (n=4).

Table 5 shows the intercorrelations of these 5 latent misconceptions, separately by problem. For both problems, misconception B1 has a moderate positive correlation with

B4, and B3 with B5. However, there is a moderate negative correlation of B1 with B2, and of B3 with B4, which may indicate problems distinguishing these misconceptions.

Table 5: Intercorrelations between diagnosed misconceptions in the two problems using the 5-bug network (above the diagonal = Problem 1; below the diagonal = Problem 2).

	Bug 1	Bug 2	Bug 3	Bug 4	Bug 5
Bug 1	--	-.19	.06	.21	.00
Bug 2	-.12	--	.10	-.01	-.01
Bug 3	.05.	.10	--	-.17	.15
Bug 4	.20	-.05	-.20	--	-.01
Bug 5	-.08	.00	.13	.10	--

We performed k-means clustering to shed light on the distributions and intercorrelation of these misconceptions and biases. For each participant, the input to the cluster analysis was the six posterior probabilities for the five misconceptions measured across the two isomorphic problems. Table 6 shows the final cluster centroids for the resulting 2-cluster solution. Cluster 1 (N=91) consists of those participants diagnosed with a moderately high posterior probability of having misconception B2, the LAN-as-path error, and B4, the infer-bridge error. Note that B4 can be viewed as a location-specific "specialization" of B2, in the sense that it has the same symptom, which is interpolation of a node into a path within a LAN. Cluster 2 (N=37) consists of participants diagnosed with a relatively high posterior probability of having all the misconceptions and biases, but especially with B3 and B2, the misconceptions with higher base rates.

Discussion

When given the task of generating all the shortest paths in a diagrammed information system, participants made many errors. Knowing the origins of the errors is important for both instruction and diagram design. Determining the origins of the errors is not straightforward, because many are ambiguous with respect to cause.

Here we followed an abductive method, based on Bayesian inference, for diagnosing the misconceptions underlying path errors. Specific misconceptions were hypothesized to explain individual surface errors and used to construct causal models represented as Bayesian networks. We select among members of a set of nested candidate models using model-fit statistics. This is in effect a form of structure learning. The model-fit statistics BIC and AIC supported the usefulness of all five bugs in affecting performance in the paths task. However, the advantage of the 5-bug solution over the 4- and 3-bug solutions is not large. By using two isomorphic diagrams in the test, the estimated models from two diagrams could be compared to assess reliability of the diagnosis.

The misconceptions that were reliably diagnosed in all models suggest the following conclusions about understanding and performance in the paths task. A small group of participants did not order objects correctly; they appeared to regard the diagrams as sets of objects with no relations among them, that is, categorically. In other words, they were able to make inferences from the nodes, but were unable to make inferences from the paths between the nodes. A larger group of participants failed to understand the LAN straight-line convention in which sets of components are completely connected. That group seemed to assume that any line could order objects. Some participants failed to make correct inferences regarding the role of bridge nodes in the diagrams. Another group processed the diagram in reading order, left to right, and failed to generate certain paths, mostly backwards ones.

All of these errors seem to arise from "reading" the diagrams incorrectly. Participants who failed to correctly order the objects seemed able to "read" or understand the nodes, but not the paths. Most participants, however, did understand both the nodes and the paths. Many did not understand that all components connected on a straight line are interconnected so that a shortest path need "visit" only the endpoints. This misconception is encouraged by the straight line that connects the components. Participants who omitted some reverse paths seemed to be making inferences from the network in reading order, but had no visual indication or feedback for each step.

This last finding is analogous to studies that show that mathematical formulae are also processed in a sequential manner (Landy & Goldstone, 2007). More generally, this work may be related to studies that claim that people infer structure from data, and then use those structures as priors (Kemp & Tenenbaum, 2008). In our study, it is possible that each participant is choosing a prior model, a structure such as a sequential chain, to understand the diagram and make inferences. If the structure is incorrect (a misconception), then the participant makes errors. In this way of thinking, our model seeks to classify each participant by discovering this unobservable structure.

The findings have general implications for education and design, since networks are so widely used to represent information. Although the mapping of elements and relations in the world to elements and relations in diagrams facilitates reasoning and inference by providing visuospatial cues, visuospatial cues or their absence can also have deleterious effects on people's interpretations of diagrams.

Once the origins of errors are known, interventions can be designed to reduce them. We have shown that at least some of the misconceptions defined above can be corrected through a simple educational intervention (Corter et al., 2008). By looking at multiple types of diagrams, it might be possible to diagnose misconceptions that apply across domains. For example, sequential biases have been shown in a variety of domains (e. g., Taylor & Tversky, 1992). It remains to be seen if simple interventions in one domain will yield returns across other domains.

Acknowledgements

The authors gratefully acknowledge the support of the National Science Foundation (awards IIS-0725223 and REC-0440103) and the Stanford Regional Visualization and Analysis Center.

References

Andersen, S. K., Olesen, K. G., Jensen, F. V., & Jensen, F. (1989). HUGIN – A shell for building Bayesian belief universes for expert systems. In N. S. Sridharan (Ed.), Proceedings of the 11th International Joint Conference on Artificial Intelligence (pp.1080-1085). San Mateo, CA: Morgan Kaufmann.

Beland, A., & Mislevy, R. J. (1996). Probability-based inference in a domain of proportional reasoning tasks. *Journal of Educational Measurement, 33*(1), 3-27.

Conati, C., Gertner, A., & Vanlehn, K. (2002). Using Bayesian networks to manage uncertainty in student modeling. *User Modeling and User-Adapted Interactions, 12*(4), 371-417.

Corter, J. E., Nickerson, J. V., Tversky, B., Zahner, D., and Rho, Y. J. (2008). Using diagrams to design information systems. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society*, (pp. 2259–2264). Austin, TX: Cognitive Science Society.

Jensen, F. V. (1996). *An introduction to Bayesian networks.* New York: Springer.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences, 105(31)*, 10687-10692.

Korb, K. B., & Nicholson, A. E. (2004). *Bayesian Artificial Intelligence.* London: CRC Press.

Landy, D., & Goldstone, R. L. The alignment of order and space in arithmetic computation", in *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 2007, pp. 382-387.

Lee, J. (2003). Diagnosis of bugs in multi-column subtraction using Bayesian networks. Unpublished Ph.D. dissertation, Columbia University.

Lee, J. & Corter, J. E. (2009). Diagnosis of subtraction bugs using Bayesian networks. Under editorial review.

Martin, J., & vanLehn, K. (1995). Student assessment using Bayesian nets. *International Journal of Human-Computer Studies, 42,* 575-591.

Pardos, Z. A., Heffernan, N. T., Anderson, B., & Heffernan, C. L. (2007). The effect of model granularity on student performance prediction using Bayesian Networks. In Conati, C., McCoy, K., & Paliouras, G. (Eds.), *User Modeling 2007: Proceedings of the 11th International Conference*, UM 2007, Corfu, Greece, July 25-29, 2007. Berlin: Springer.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems.* San Mateo, CA: Morgan Kaufmann.

Taylor, H. A. & Tversky, B. (1992). Descriptions and depictions of environments, *Memory and Cognition*, *20,* 483–496.

Tversky, B. (2001). Spatial schemas in depictions. In M. Gattis (Editor), *Spatial schemas and abstract thought.* (pp. 79-111). Cambridge: MIT Press.

Tversky, B, Zacks, J., Lee, P. U., & Heiser, J. (2000). Lines, blobs, crosses, and arrows: Diagrammatic communication with schematic figures. In M. Anderson, P. Cheng, and V. Haarslev (Editors). *Theory and application of diagrams.* Pp. 221-230. Berlin: Springer.

Spiegelhalter, D. J., Dawid, A. P., Lauritzen, S. L., & Cowell, R. G. (1993). Bayesian analysis in expert systems. *Statistical Science*, *8*(3), 219-247.

VanLehn, K. (1990). *Mind bugs: the origins of procedural misconceptions.* Cambridge, MA: The MIT Press.

VanLehn, K., & Martin, J. (1998). Evaluation of an assessment system based on Bayesian student modeling. *International Journal of Artificial Intelligence in Education*, *8*(2), 179-221.

Williamson, D. M., Almond, R., Mislevy, R. J., & Levy (2006). An application of Bayesian Networks in automated scoring of computerized simulation tasks. In D. M. Williamson, R. J. Mislevy, and I. I. Bejar (Eds.), *Automated Scoring of Complex Tasks in Computer-Based Testing*, pages 201–257. Hillsdale, NJ: Lawrence Erlbaum.