1-1-2006

Methodological challenges for identifying and coding diverse knowledge elements in interview data

Victor R. Lee
Utah State University

Moshe Krakowski
Yeshiva University

Bruce Sherin
Northwestern University

Megan Bang
University of Washington - Seattle Campus

Gregory Dam
Northwestern University

Methodological challenges for identifying and coding diverse knowledge elements in interview data

Victor Lee, Moshe Krakowski, Bruce Sherin, Megan Bang & Gregory Dam
School of Education and Social Policy
Northwestern University

Introduction

As discussed in the earlier papers in this symposium (Krakowski *et al. this session;* Sherin *et al. this session*), we take the position that the science education community has a lot to gain from examining the clinical interview as a setting for conceptual change, albeit at a small scale. On the one hand, there is the potential for increased understanding of the mechanisms of cognitive change when we conduct in-depth analyses of such change that occurs at shorter, more manageable time scales. We may, for example, be better able to localize specific interactions that promote certain forms of conceptual change or catalogue regular patterns of student reasoning around complex subject matter. On the other hand, there is the potential to better understand the instrument itself and how it can be best used to measure knowledge and thinking.

The approach adopted here differs, in some important respects, from the predominant approach in the science education community. In the dominant approach, the clinical interview is used as a means for reading out students' thinking; in essence, several minutes in a clinical interview are collapsed to a single data point, which can then be more easily categorized and grouped with other similar data points. One version of this takes the form of what diSessa, et al. (2004) referred to as *model-mapping*, in which interviews with individual students are coded in terms of internally coherent mental models of physical phenomena (e.g., Ioannides & Vosniadou, 2002; Samarapungavan & Wiers, 1997; Vosniadou & Brewer, 1992).

We take the position in this symposium that there is a different, and theoretically more productive way to go. We believe that the earlier model-mapping work is productive, and can be built upon. But, we also believe that, at the present time, the research community can benefit from an increased attention to what is happening in a clinical interview – unpacking the process by which coherences emerge. In the best case, we would like to know in much more detail what basic knowledge elements are activated, when those elements are activated in response to changes in an interview interaction, and the subsequent role that those elements play in the activation of later knowledge. This is our ideal. However, moving from raw video of clinical interviews to something approximating this fine-grained account is a significant undertaking, one that goes beyond the state of the art in the field.

Our initial goal was to conduct the analysis using the "mode-node" framework presented in Krakowski et al. (this session). However, we found that any efforts to operationalize this framework in a manner that allowed rigorous and systematic coding of the interview data ran us on a collision course with one or more major modeling challenges. As we documented and examined these challenges during the course of our analysis, we began to realize just how pervasive these challenges were to our modeling work and how they are often unavoidable in any effort to model knowledge and cognition.

Describing these challenges is the focus of this paper. In future papers, we plan to examine in further depth how these challenges are manifested throughout other work on knowledge representation and modeling cognitive processes. For the purposes of this paper, though, we take some more modest first steps by enumerating and describing these challenges in the context of our own analysis of interview data. We will present these challenges in two sets, one pertaining to the establishment of basic knowledge elements that fit our modeling vocabulary – the nodes –

and the other having more to do with the application of these elements to code specific interviewing episodes.

Some of the difficulties that we face are simply endemic to the task of *modeling*. In all instances of modeling, whether they involve capturing the dynamics of a physical phenomenon or a class of human cognitions and behaviors, there are approximations that must be made. We must choose what aspects to represent, whether or not to recognize each entity as distinct from other entities, and to account for the sorts of behaviors and interactions that each entity will have. However, we believe that there are noteworthy manifestations of these sorts of approximations that emerge, that are specific to the task of modeling thinking and reasoning of students in clinical interviews. It is for that reason that we believe that the endeavor undertaken in this paper is worthwhile.

The Major Methodological Challenges

Recall that, in the second paper of this symposium, we used the term "node" to refer to any kind of element in the mental ecology. This theoretical construct was intended to encompass knowledge elements of multiple types and at multiple levels of abstraction. They may be, for example, mental models (e.g., the shape of the Earth), conceptual schemas (e.g., *dying away*), fact-like commitments (e.g. "It's summer in South America when it's winter in North America"), or unfamiliar lexical items (e.g., the word "tilt" as it relates to the Earth's orbit).

Describing all of these elements as nodes provides some analytic advantages. Most notably it allows us to be encompassing in the kinds of knowledge that we consider in our analysis; we can include a great diversity of kinds of knowledge, without having to be specific about differences in the form or other specific properties of that knowledge.

But this same flexibility also exacerbates some methodological challenges. As described in the second paper, we have attempted to code our interview data in terms of a specific set of nodes. The problem is that, in treating all knowledge as nodes, we have lost some of the constraints that might help in identifying specific nodes. To illustrate, consider the following two pre-interview responses from students who were asked "Why is it warmer in the summer and colder in the winter?"

<u>Angela</u>

A1. Angela	That's because of the sun is in the center and the Earth moves around the sun and the Earth is like at one point in the winter, it's like farther away from the sun and towards the summer it's closer it's near, towards the sun.

<u>Mark</u>

M1. Mark	Part because of the Earth's axis, uh, revolution around the sun. It hits an angle, say... it's the Earth's axis, and sometimes the Earth, <gestures> this is the sun and this is the earth and it's like slanted this way <tilts hand> so it takes longer for it to reach and the sun's UV rays aren't as strong and it's also at an angle, it slants.
M2. Interv.	Why is it not as strong?

M3. Mark: Cause it takes longer to travel. It's weaker.

As discussed in the first paper of this symposium, these two students, Angela and Mark, gave very different explanations of the seasons. Angela's explanation is a "closer-farther" explanation: The earth orbits in an ellipse, and summer occurs when the Earth is closer to the sun, while winter occurs when it's farther away. Mark's explanation, in contrast, is a minor adaptation of the accepted scientific explanation. According to Mark, the Earth is tilted. This tilt causes one hemisphere to be closer to the sun – thus experiencing summer – and one hemisphere to be tilted away from the sun.

Though these explanations are very different, they clearly share some elements. Both students talk about the Earth's orbit around the sun. Angela says "the Earth moves around the sun" and Mark says "because of the Earth's axis, uh, revolution around the sun." Almost certainly, we'd want to code for similar elements of knowledge in these two cases. But not all cases are this clear. Both students' explanations apparently involve some reference to changing distance of all or part of the Earth from the sun. Angela says "it's like farther away from the sun and towards the summer it's closer it's near, towards the sun." But notice that Mark doesn't explicitly refer to the Earth's distance from the sun. Instead, he says that the sun's rays will be weaker because it takes longer for them to reach the Earth. The difficult question is: Can we code both explanations as involving the same element? For example, is there some shared understanding like "sources have weaker affects farther from the source?" Or are there meaningful differences that we should capture?

This and other questions emerged over the course of our empirical work. We have found it helpful to group and characterize these questions as seven specific recurring problems which we have encountered. These problems are associated with two kinds of empirical activities: the generation of an inventory of nodes for use in coding interview data (node generation) and the application of those codes to specific interview episodes (node application). In the sections that follow, we will discuss these specific problems. The node generation problems include: (1) the *completeness* problem, (2) the *parametric* problem, (3) the *assembly* problem, (4) the *abstraction* problem, and (5) the *hairsplitting* problem. The node application problems include: (1) the *inference* problem and (2) the *carryover* problem.

Node Generation Problems

Recall in the second paper that our empirical work involved first generating an inventory of nodes, ultimately resulting in a coding scheme consisting of 109 nodes. As we constructed and refined our inventory, we gained increased familiarity with five recurring problems. These five problems are all, to a degree, related to each other, and thus some overlap should be expected.

The Completeness Problem

The first problem we discuss is the completeness problem. In some respects, this problem is the root cause for all of the other problems we will discuss. As mentioned in Sherin, et al. (this session), our approach requires us to account for much more of the students' knowledge than has traditionally been done in the past. If we wish to be complete, the amount of knowledge that must be modeled is potentially enormous. The main question here is: how can we reasonably account for such a large amount of knowledge in our model?

As an example, consider this interview excerpt from a middle school student, Kurt. In Kurt's initial explanation of the season, he attributed the seasons to the Earth's rotation, and he stated that the side of the Earth facing the sun experiences summer, while the side facing away experiences winter. The interviewer then followed up with one of our standard prompts, asking "if it's winter in Chicago, where is it summer?" When Kurt appeared to struggle, the interviewer offers Australia as a possibility.

1.	Interv.	So if it's winter in Chicago, where is it summer?
2.	Kurt	Probably in...kinda in Antarctica? No, not Antarctica...
3.	Interv.	Like Australia?
4.	Kurt	Yeah. Kinda downward, tilting up, like how it's.... I was trying to think. I wouldn't say Europe. Cause Europe also gets cold.
5.	Interv.	What gets cold?
6.	Kurt	Cause I wouldn't say Europe cause it's kind of like on the same, kind of on the same even but downward cause it kind of tilts the Earth, all the countries that are down here, like Australia, would go up.

What parts of Kurt's BML are active in this segment? Given that he is already talking about the seasons, we can confidently assert, based on his acceptance of the interviewer's question and response (Line 2), that some knowledge of the seasons is involved. Included in that is knowledge that summer is warm and winter is cold, and different seasons may co-occur at different parts of the Earth. By Line 4, we can also confidently assert that Kurt knows that Antarctica and Australia are of the same scale and type (he didn't name a specific city or an Ocean). In addition, since he says "Kinda downward, tilting up" and is gesturing to indicate the movement of the Earth, he appears to have activated knowledge of Australia's relative location on the Earth as being downward, though because of some tilting motion, it can be imagined as being in a different location that may be more conducive to experiencing summer. Moments later, in the same line, he mentions Europe. At this point, we are confident that his reasoning now involves knowledge about Europe, also being of the same scale and type as Antarctica and Australia, and that it is a place that gets cold.

Then, at Line 6, Kurt seems to have activated new knowledge, something that allows him to interpret Europe and the United States (including Chicago) as being on the same latitude. The tilted Earth is activated (though we can ask if the way that tilt is understood here was the same as in Line 4), as well as some knowledge of the effect of tilt on relative position of static entities on a surface.

The problem in analyzing Kurt's interview is simply the obvious one: There is so much knowledge that is potentially active and relevant, that it is very difficulty to truly model it all. Our response to this problem has in part leveraged the generic quality of nodes. As discussed in the second paper in this symposium, our node scheme includes what we called *composite* nodes – nodes that essentially encompass a potentially large set of cognitive elements – but whose structure was not further analyzed. For example, we believed that we could be quite certain that the students we interviewed know a great deal about the seasons – what times of year they tend to happen, what each season feels like, etc. Rather than enumerating all the many aspects of this knowledge, we collapse them into a "basic knowledge cluster."

While the inclusion of basic clusters in our scheme does not fully resolve the challenges raised by the completeness problem, we believe it is a step in a positive direction. It allows us to make progress, while acknowledging that there is knowledge that would have to be included in an analysis that was more fully in alignment with our completeness principle. We might never be able to avoid making approximations of this sort. The challenge for future work is to develop criteria that allow us to make these approximations in a less ad hoc manner.

The Parametric Problem

As we enumerated our inventory of nodes, we needed to ask ourselves "what features of knowledge elements are we most interested in capturing?" We refer to this as a parametric problem. Some approaches that we considered as we generated an inventory of nodes:

- Focus mostly on the *representational form* of the knowledge, whether it takes the form of a conceptual schema, a fact-like commitment, a mental model, or something else.
- Take a more *semantically-driven* approach that focuses on the *content* of any knowledge element. This may be productive if we want to establish some sort of metric that compares the student's conceptual ecology to an expert's or some other target.
- Adopt a *genetic* perspective that is concerned with the origins of the knowledge – where, when, and how the knowledge is acquired, both within and prior to the interview.
- Describe knowledge on a basis that is more local to the clinical interview, such as *who – interviewer or student – contributed a node* to the conceptual common ground.

In our own work, we have prioritized content and form as parameters. When analyzing an interview episode, we focus, first, on what the knowledge is about, as inferred by the contents of the students' responses, and second on the form. This choice is motivated by a simple consideration: Our primary concern is with modeling the reasoning of the student, as it occurs during the interview. Truly constructing a knowledge *requires* that we attend to the form and content of a student's knowledge.

The Assembly Problem

As mentioned earlier in this paper, our move to include both simple and composite nodes has introduced some additional complications. One of those complications is that we must decide when to treat something as a collection of simple nodes, or as a single composite node. Consider an excerpt from an interview with Ali.

1.	Interv.	The question is why is it warmer in the summer and colder in the winter? That's what I want to know.
2.	Ali	Because of the su/ because of the Earth's movement… er...yeah, the axis of the Earth.
3.	Interv.	So how exactly does that work? How does it make it...?
4.	Ali	Well, like when the Earth goes around the sun it's kind of tilted. It turns too like every 24 hours it turns and so that changes our seasons

5. Interv.	But how does//
6. Ali	So when it's away from the sun it's colder and when it's like directly in front of the sun it's hot

In Line 4, we could reasonably interpret what Ali says as being suggestive of one node that generates the reasoning (a spatial mental model of the Earth's movement that directly accounts for the seasons). However, we could also reasonably interpret this as being the product of six distinct nodes (Earth moves around the sun, Earth is tilted, Earth turns, 24 hours associated with Earth's movement, basic seasons cluster, movement of bodies can cause changes).

This problem also goes in the opposite direction. Some nodes that appear to be simple could potentially be decomposed further. For example, when Ali says in Line 6 "when it's away from the sun it's colder and when it's like directly in front of the sun it's hot," there could be very different configurations of nodes at work. One could be that she has activated a content-specific node "towards the sun yields more heat than away from the sun" or she could have activated an assembly of simple nodes such as "towards a source yields more effect than away from the source" and "the sun is the source of heat." Differentiating between those poses a problem for generating an inventory of nodes. There is an inevitable struggle to find the appropriate level of description that may best approximate how that knowledge is encoded mentally for the student.

The assembly problem is not one that we see as specific to the work we do. For example, in theories about cognitive schemata (e.g., Rumelhart, 1980), the same issues arise when trying to identify what can be decomposed further and what should be considered as a single unit. Almost any schema could be broken down into smaller pieces and at any given moment, the whole or some piece may be active. If we want to model this precisely, then when do we hit bottom?

The coding scheme we have developed provides some means for us to describe some types of assemblies. For example, as described in the second paper in this symposium, it includes mental models, as well as what we called "basic clusters."

The Abstraction Problem

Another related problem involves the level of abstraction or specificity of knowledge. Consider the same Ali excerpt discussed above. It appears that at the moment of Line 2's utterance, Ali is still formulating her answer. She is somewhat hesitant in her response, interrupting herself, and adding additional facts (such as the "axis" in Line 4) as she is formulating a fuller answer. By Line 6, she has a more detailed explanation, and we can catch a glimpse of how her thinking unfolded. She seemed to start with a general idea of movement involving the Earth that later was specified as movement around the sun and also movement around an axis.

In assigning codes to each line, what should we consider her understanding at the moment of Line 2 to be? Is her understanding of movement at that moment the same as what comes out in Line 6? Note that, in Line 2, she only refers, quite generally, to "the Earth's movement." In contrast, in Line 6, she says that the earth "turns…every 24 hours."

But if we were to do so, we would be likely to overlook a very important aspect of Ali's statement in Line 2, that she has at that point already given a form of coherent explanation. Ali's explanation at that point involves a causal agent (Earth's movement) and a resultant effect (the phenomena of the seasons), though it is rather underspecified. Because we wanted to capture the various kinds of coherences that emerge at multiple levels of abstraction, it seemed that building something into our analysis that could capture this would be the appropriate move. Also, there are some situations where we could reasonably expect a student's knowledge to be the level of generality of "Earth moves" and not know exactly how it moves.

The Hairsplitting Problem

The final problem we discuss in this section deals with the challenges involved in distinguishing nodes that are very similar in content. Recall the earlier pair of examples from the interviews with Angela and Mark. Both of these students produced explanations for the seasons that referred to the distance from the sun.

Angela

A1. Angela That's because of the sun is in the center and the Earth moves around the sun and the Earth is like at one point in the winter, it's like farther away from the sun and towards the summer it's closer it's near, towards the sun.

Mark

M1. Mark Part because of the Earth's axis, uh, revolution around the sun. It hits an angle, say... it's the Earth's axis, and sometimes the Earth, <gestures> this is the sun and this is the earth and it's like slanted this way <tilts hand> so it takes longer for it to reach and the sun's UV rays aren't as strong and it's also at an angle, it slants.

M2. Interv. Why is it not as strong?

M3. Mark: Cause it takes longer to travel. It's weaker.

The big question here is: how similar or different are the general reasoning nodes in these two situations? As discussed earlier, both students' explanations apparently involve some reference to changing distance of all or part of the Earth from the sun. Angela focuses directly on the distance from the Earth to the sun. She says "it's like farther away from the sun and towards the summer it's closer it's near, towards the sun," but she gives no explanation of why this might be the case. In particular, she makes no mention of anything that might be traveling between the sun and the Earth. Mark, in contrast, focuses on the sun's rays, which travel from the Sun to the Earth. Furthermore, he doesn't explicitly mention the changing distance between the Earth and sun. Instead, he appears to focus on the *time* it takes the rays to reach the Earth, saying that they will be weaker if they take longer to travel. The question here is whether there is any real difference in Mark and Angela's explanations. For example, is it at all consequential that Mark focuses on the time it takes for the rays to travel, rather than the distance they might travel? Does it matter that Angela has not explicitly mentioned the entities that are traveling between the sun and Earth?

The differences described here are subtle ones. Determining how much those differences matter is the heart of the hairsplitting problem. To some extent, the issue here is a pragmatic one. The more that fine distinctions are introduced into the coding scheme, the more difficult the scheme is to use, and the harder it is to obtain reliability. However, if we tend too much to approximation and consolidation, then we may miss important nuance – nuance that might be important for understanding the reasoning we observe.

The Node Application Class

The second class of methodological challenges involve problems that are of a slightly different nature than those described earlier. These can, roughly speaking, be thought of as occurring after our node inventory is created, when the major task at hand is to apply an accepted set of nodes to data. The two problems we discuss here are the *inference* problem and the *carryover* problem.

The Inference Problem

Confronting the inference problem is inevitable when working with clinical interviews. Unless strictly computational tools are used (such as the one described in the next paper), we must rely on our own knowledge of the language used in the interaction to make interpretations. In many respects, this is not so problematic. As participants in a particular cultural group shared with research subjects, we as humans are quite effective at inferring the meaning of a spoken utterance or understanding the general flow of conversation in a conversational interaction. If we did not have these competencies, such things as reliability in the analysis of verbal data would be impossible to obtain, and consensus in interpretation could never be reached.

Nonetheless, in attempting to code interviews in terms of nodes, we frequently must wrestle with exactly what can be inferred from a student's utterances. For example, suppose that a student says "the seasons are by the changing distance of the Earth from the sun." What can we infer about the sun? Can we infer that the student believes that the sun is the source of heat that warms the Earth? Does it warm the Earth by providing light or perhaps some other unspecified essence?

In coding Angela excerpts discussed above, we face exactly this problem when Angela says "Earth is like at one point in the winter, it's like farther away from the sun and towards the summer it's closer it's near, towards the sun." As discussed in the hairsplitting problem, we must make a decision whether to infer she has active knowledge of the radiation emitted from the sun and how that relates to the different seasons. Also, we must make inferences about what is in her mental model of the Earth's movement and what sort of entity she understands the Earth to be.

The Carryover Problem

The last major challenge that we discuss in this paper involves another inherent uncertainty involved in this sort of microanalysis of interview data. Recall that as part of our framework, when a set of nodes has been activated, we expect that from within that set, different constructions may emerge given the changing contextual conditions of the interview. Along with these reconfigurations of existing nodes, new nodes that were not yet active gain prominence in subsequent constructions. Using the conservative approach of coding-when-spoken mentioned above, identifying the newly active nodes can be made a bit more manageable. However, keeping track of nodes that were active earlier introduces another methodological problem. For the purposes of coding interview data, how would we best track or code nodes that we were certain were active before but are not explicitly mentioned later?

For example, students often mentioned early on in their interviews that the sun is the source of the heat on Earth. However, this fact is usually not mentioned again as the interview progressed, likely because as a communicative interaction, that idea has been established as part of a shared common ground of information between the interviewer and interviewee (Stalnaker, 1978). In everyday communication, it would seem needlessly redundant if the idea that the sun is the source of heat on Earth is mentioned repeatedly, and violate conventions that we would normally follow (Grice, 1975).

This is further complicated because, for the most part, we can assume in everyday interactions that what enters the common ground stays there for the duration of the interaction. However, this is not a safe assumption for interview situations. Examples discussed in Sherin's paper (this session) illustrate how ideas that were accepted as part of the common ground at one point are often later rejected when new explanations emerge. Sometimes, this shift is explicit and involves the student making an outright rejection of what had been said before. However, there are often cases where this shift is gradual, as ideas in the common ground are further backgrounded over time and eventually are no longer a part of the active DMC.

The challenge in the carryover problem is determining when nodes are carried over, both from utterance to utterance and from construction to construction. Coding the entire set of nodes that are active, even those that are not spoken, at every line of utterance would violate one of our earlier strategies – to code nodes as they are reflected in spoken utterances – and would also be very inefficient. We could also expect both validity and reliability to drop significantly as such a coding strategy would reintroduce instantiations of the inference, assembly, completeness, and hairsplitting problems and create obstacles to reaching consensus.

An alternative that we have attempted to implement is to allow for some partial segmentation of interviews based on topic or task shifts. Because interviews involve a set of roles in which the interviewer generally sets the agenda for discussion and the interviewee responds, we have attempted to assign codes to interviewer utterances as specific kinds of speech acts (Austin, 1962) specific to clinical interviews. There were a few interviewer speech acts that we found particularly helpful. One was the introduction of a core prompt, in which the interviewer strictly followed the protocol by giving the students a general task such as "Explain why it's warmer in the summer or colder in the winter." Locating these in the interviews has enabled us to do a coarse segmenting of the data, and for utterances immediately following core prompts, we attempt to do an exhaustive coding that can be used as a reference point for later utterances when some node was not spoken but seems to be active because it is assumed to be in the common ground.

Second, we have made a point to monitor interviewer acts in which the interviewer provides backchannel feedback (Duncan, 1974). When these transcriptions were made, we were careful to include as much as possible about any verbal utterances made by the interviewer, including, for example, both ratified turns and interruptions. There are many occasions in the raw transcripts in which a student has paused, sometimes indicating that they have ended their speaking turn and sometimes expressing their own uncertainty about whether or not to continue. In both of these situations, one standard interviewer act was to decline his or her own turn at that moment by saying something in agreement such as "I see" or "Oh, Okay" "mm hmm" and thus passively encouraging the student to continue. Regardless of whether or not it was a single turn or many, labeling and filtering out these instances of backchannel feedback from the interviewer has allowed us to collapse segments of the interview into extended utterances. These extended utterances sometimes span multiple lines of student utterances, but provide a window from which we can assume that nodes are carried over to later utterances even though they have not been mentioned explicitly for an extended period of time.

Third, we have made attempts to explicitly code interviewer-contributed nodes. The assumption is that explicit mention by the interviewer reintroduces or highlights a node that was at some point earlier part of the common ground, and thus makes it manifest in the student's thinking at the time. That gives us some basis for assuming with higher confidence that a node not explicitly mentioned by the student is still active.

These innovations have been useful for us and we take them to be suggestive of ways in which conceptual change research may be able to adapt some ideas from linguistics and studies of discourse to make some progress. Regardless, these tools do not work in all occasions, and we find that we sometimes must take our best guesses on some nodes and in any formal presentation, do our best to make our rationale as explicit as possible and make clear the implications of other potential interpretations that could be made.

Conclusion

The purpose of the larger symposium, of which this paper is a part, is to discuss approaches and challenges for using clinical interviews as a tool for capturing science knowledge. In this paper, we narrowed our attention to a focus on the *challenges*. To some extent, the challenges we identified are unique to our research program. As we have discussed repeatedly, we believe our program is unique in the extent to which it attempts to be complete and encompassing in modeling the unfolding events in a clinical interview. However, to a certain extent, we believe that the problems discussed here are universal; they apply to any attempts to understand the knowledge evidenced by students in clinical interviews or, indeed, to any attempt to see knowledge in action.

In our discussion, we divided the challenges we faced into 7 problems. But a number of themes reappeared across all of these problems. First, we were repeatedly faced with the need to mitigate the problems posed by our own desire for completeness. In truth, it is not possible for us to produce an analysis that is "complete" in the way demanded by the principles presented in the first paper in this symposium. This means that we have needed to develop strategies for ignoring some features of an interview – for ignoring some knowledge, and collapsing collections of knowledge into a manageable form. Second, we needed to wrestle, in a number of ways, with where and how to draw distinctions, and where to treat elements of knowledge as the same. Here, again, there was a tension with pragmatics. Ignoring distinctions makes it easier to code, and to code reliably. But in ignoring these distinctions, we may be ignoring nuance that is critical for understanding a particular interview.

In the end, we have not come close to offering real solutions to these problems. Indeed, it may not even be clear from these discussions that there *are* solutions to these problems, and that the program we have laid out is practical. Nonetheless, we hope to have, in this paper, at least begun to map out some of the challenges that we believe must be faced, if we want to use clinical interviews as tools for getting at the science knowledge of students.

References

Austin, J. L. (1962). *How to do things with words*: Oxford University Press.

Dam, G., & Kaufman, S. (2006). Using latent semantic analysis for interpreting clinical interview data, *2006 Annual Meeting of the American Educational Research Association*. San Francisco, CA.

diSessa, A. (1987). Toward an epistemology of physics. *48*.

diSessa, A. A., Gillespie, N. M., & Esterly, J. B. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science, 28*, 843-900.

Duncan, S. (1974). Some signals and rules for taking speaking terns in conversations. In Weitz (Ed.), *Nonverbal communication*: Oxford University Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 41-58). New York: Academic Press.

Ioannides, C., & Vosniadou, S. (2002). The changing meaning of force. *Cognitive Science Quarterly, 2*, 5-61.

Krakowski, M., Lee, V., Sherin, B., Bang, M., & Dam, G. (2006). Modes and nodes: A cognitive framework for capturing conceptual dynamics, *2006 Annual Meeting of the American Educational Research Association*. San Francisco, CA.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33-58). Hillsdale, NJ: Lawrence Erlbaum.

Samarapungavan, A., & Wiers, R. W. (1997). Children's thoughts on the origin of species: A study of explanatory coherence. *Cognitive Science, 21*(2), 147-177.

Sherin, B., Krakowski, M., Lee, V., Bang, M., & Dam, G. (2006). Conceptual dynamics in clinical interviews: An introduction, *2006 Annual Meeting of the American Educational Research Association*. San Francisco, CA.

Stalnaker, R. C. (1978). Assertion. In P. Cole (Ed.), *Syntax and sematics: Pragmatics 9* (pp. 315-332). New York: Academic.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology, 24*(4), 535-585.