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

Wilensky U. & Resnick M. (1999) Thinking in levels: A dynamic systems approach to making sense of the world. Journal of Science Education and Technology 8(1): 3–19.

Yu (Bryan) Guo's research interest is in using complex-systems approaches and computer-based models to help people understand complex and controversial social phenomena. His dissertation investigates high-school students' learning processes and outcomes about wealth inequality in "Mind the Gap" – a computer-based participatory simulation curricular unit that he designed. He received his PhD in learning sciences from Northwestern University.

Received: 28 May 2019
Accepted: 10 June 2019

Authors' Response
New Questions About New Methods in Old Contexts

Arthur Hjorth
Aarhus University, Denmark
arthur/at/cs.au.dk

Uri Wilensky
Northwestern University, USA
uri/at/northwestern.edu

> Abstract • Designing, implementing and assessing the effects of classroom-based learning experiences spans across a wide variety of methodological, epistemological and design-related issues. Additionally, the use of data mining and computational methods for supporting qualitative data analyses is still new to the field. Potentially because of this, we received ten good, but quite different commentary questions, which we have organized under five headings. In this response, we address each of them to provide a more thorough background or reasoning behind the decisions we made in our target article.

« 1 » We wish to thank the four commentators, who asked a total of ten interesting and thought-provoking questions about our article. In our response, we have categorized these questions under five headings.

Choice of theory and approach

« 2 » **Jim Clayson** (Q1) raises a good question that we may have assumed the answer to, without making a case for it in our target article, i.e., why Knowledge-in-Pieces (KiP) can be applied in the social sciences as well as in math and physics.

« 3 » First, a bit of context to this question. Knowledge-in-pieces originated with Andrea diSessa's work on physics students' reasoning at college level. Andrea diSesssa, and others, haven taken KiP approaches to understanding how students reason about a wide variety of phenomena related to physics, such as gravity, resistance, movement and force, proper time, the shape of the Earth, Earth's four seasons, and the springiness of a spring (Andrea diSessa 1993 2002; Kapon & diSessa 2012; Levrini & diSessa 2008; Sherin, Krakowski & Lee 2012; Vosniadou & Brewer 1992). KiP has shown itself to be a powerfully generative approach to documenting and understanding students' knowledge of and conceptual change in physics. Of course, it does not necessarily follow that this framework would work outside of physics.

« 4 » However, more recently, researchers have started applying KiP to reasoning about non-STEM-related phenomena. Thomas Philip took a KiP approach to understanding the ideology underlying teachers' reasoning about students and race in his work on what he calls *ideology-in-pieces* (Philip 2011). Last year, Arthur Hjorth, one of the authors, participated in a symposium at the International Conference of the Learning Sciences 2018 called "Knowledge Analysis Outside the STEM Classroom" (Anderson et al. 2018). In the symposium, researchers presented their work on using Knowledge Analysis, which posits itself as the methodological approach associated with KiP (diSessa, Sherin & Levin 2016), to study reasoning about social issues ranging from restorative justice, and teachers' reasoning about pedagogy, to ethical sensemaking in engineering. While much of this work is still exploring the potential limits of using KiP outside of STEM, at this point a reasonable fit has been established by this emerging body of literature.

Data analysis and analytical methods

« 5 » **Clayson** emphasizes our statement that we used "a pragmatic and somewhat promiscuous approach to identifying the individual knowledge pieces" (§30), and asks us to explain this, and give a better account of our analytical approach and methods, and why we chose to focus our attention on student responses to the question that we did focus on, and not on one of the four other questions.

« 6 » First, a brief addendum to the phrase that **Clayson** quotes, and which may have not communicated what we intended. A central contribution of diSessa's work with KiP was the identification of *phenemonological primitives* (or "p-prims") that guide students' intuitive sensemaking of physics. The full definition of p-prims is long to recount here, but two aspects are important to address: First, their origin matters. They must have come about through the *everyday interactions* that people have with the physical world. Second, their size matters. They must be the *minimal* parts of an explanation at which students bottom-out, and for which there is no answer to "why do you think that?" When we say that our identification of causal nodes was promiscuous, we mean that we did not discriminate by origin or by size of the apparent underlying causal reasoning knowledge piece.

« 7 » Our primary goal during the qualitative coding phase of our analysis therefore aimed to identify common causal claims across student responses and to uncover patterns in how they are co-activated. As mentioned in the target article (§28), we asked students five different questions. The first three focussed on the relationship between commute time and income, the fourth asked students to reason about income and access to parks and leisure areas, and the fifth asked students to give examples of how one can use urban planning to improve on a city (but left the definition of "improvement" open). As we also mentioned (§16), some students did not answer the last questions due to technical problems, so we chose to focus on the questions for which we had the most data, meaning the three questions relating to income and commute time. The first two of those questions were phrased to probe students' ideas of individual causal mechanisms by asking open-ended questions. The first was, "how might a person's income impact their commute time." Here, the hope was that students would list the various factors that they thought might contribute to an explanation in response to the question. The purpose of the second question was to start getting students to think *comparatively* about higher-income and lower-income people and about what might be different about these two groups, and we asked them, "How do you think commute times of low-income and high-income people compare?" Finally, we wanted students to explain a concrete phenomenon and asked them, "Can you explain why a wealthy person's income might make their commute time longer than a poor person's?" The open-ended nature of the first two questions meant that students' responses often attended to different aspects of the phenomenon, and that they were therefore not comparable. The focused nature of the third question meant that students' responses were comparable, and this is the reason we decided to focus on that in our target article.

« 8 » Finally, **Clayson** (Q2) emphasizes our labelling of Association Rule Mining (ARM) as agnostic with regard to the ontology of its data, and argues that we should have informed the reader about using Exploratory Data Analysis (EDA) or other more recent methodological approaches to look closely at the raw data before summarising our statistical conclusions.

« 9 » When we labelled ARM as agnostic with regard to the ontology of its data, we meant this strictly in a causal sense. This stands in contrast to many statistical methods in which one variable is assumed, a priori, to be causative. As we did not, a priori, want to assume that certain knowledge pieces (or "causal nodes") were more important or central to changes in students' conceptual change, we chose ARM because it does not make any such assumptions.

« 10 » Further, and to address **Clayson**'s stated Q2, we *did*, to some extent, use EDA when we had to make decisions about which rules to label as "important." As mentioned in paragraph 45, the ARM identified 433 different association rules. As part of our process of deciding which rules to include and which to exclude, we iteratively increased the thresholds, respectively, of lift, support, and confidence, to explore how sensitive our data were to these threshold variables, and at which point the number of included rules was sufficiently reduced to make a qualitative analysis of each one possible. Once we saw rules that were all meaningfully interpretable to us, we set the thresholds stated in the target article.

Interpretations and findings

« 11 » **Bruce Sherin** raises a concern about using data-mining approaches to learning data analysis, arguing that we often can find patterns in places where they are not meaningful, and wonders in his Q1 how to avoid the perils of prolific pattern-finding.

« 12 » On the completely opposite end of the spectrum to the open-ended nature of data mining is pre-defining a set of criteria for assessing whether responses are showing increasing signs of sophistication or completeness (e.g., Krist & Rosenberg 2016; Hjorth & Krist 2016). Taking this narrower approach requires a fairly well-developed understanding of learning in the area, however, and runs the risk of identifying students' reasoning as less sophisticated if it does not follow the pre-defined definition. The more exploratory approaches inherent in data mining allow us to discover new patterns in the development of sophistication in students, at the risk of what **Sherin** calls the "perils of prolific pattern-finding."

« 13 » **Sherin**, to some extent, answers his own question by suggesting that being able to tell a convincing story about findings is an important litmus test. While we intuitively agree, this answer simply shifts the task of assessing "what is a good data mining analysis?" to "what is a good story?" So, a more generalizable and replicable approach relies on sets of standards for parametric thresholds: for ARM and similar methods, our field should focus attention on finding these 'reasonable and meaningful thresholds. Almost certainly, defining *and* applying these standards will be complex, as they would need to address the challenges coming out of the variety in sample sizes, the variety in data types, the variety within students' responses, etc. that our field addresses in its non-computational approaches. However, establishing these standards, or at the very least, defining a set of heuristics that can help researchers establish some degree of reasonableness is necessary in the near future.

« 14 » Later, in his Q2, **Sherin** argues that the arrow in association rules, e.g., {a, b}→{c, d}, cannot be interpreted causally. So, he asks how these rules should be interpreted? And, in particular, what the relationship is between identified rules and the students' causal explanations.

« 15 » First, we agree that there is no causal implication of the arrow when looking at an association rule in isolation. However, it is important to note that we look at *changes* in a group of students as expressed by an association rule, i.e., changes to the rules that determine the right side of the arrow, holding the left side constant. In this view, we can interpret changes to the rules in two (related) ways: one that relates to students' causal reasoning, and one that relates to the design of the learning activity. With regard to the former, for those rules where we see a positive change, we can interpret the set of causal nodes on the left-hand side of the arrow as indicating a kind of cognitive-structural or explanation-structural readiness to integrate the causal nodes on the right-hand side of the arrow into their causal explanations. Of course, an explanation of why those causal nodes make it easier for students to integrate those other causal nodes lies outside of what an ARM can answer, and would probably require clinical interviews or a deeper analysis of students' reasoning during the activity. With regard to the latter, we can interpret the changes in ARM rules as indicating that certain causal chains or causal relations are made more salient to students in the learning activity or the model and that we see a particular emphasis on these causal relations for that reason.

Design of the activity

« 16 » **José Valente** (Q1) first argues for the importance of "debugging" when students iteratively design and construct artefacts and then asks, whether it is possible to record the revisions students made during model simulation so that the analysis of these changes can contribute to an understanding of how students debug their misconceptions related to the question we focus on in our §28.

« 17 » While we did not mention this in the target article, we coded the ReGrowing Chicago model (Hjorth & Wilensky 2017) specifically to support and document students' reasoning during the debugging process in a variety of ways. First, we elicit students' thinking during the process: Before students begin a new or revised city design and before the simulation runs, they are asked to explain why they designed their city the way they did, and how they think that their design decisions will affect the way in which the city will "grow." At the end of each simulation, when the city has stopped growing, students are then provided with two sources of data: spatial data in the model view and bar charts next to it, in order to assess how successful their design was in achieving the outcomes, and to reflect on whether their initial reasoning fell short. After students have submitted their reflections, the ReGrowing Chicago model is programmed to create a chronological overview of all changes and the accompanying reasoning that students have gone through in an HTML document. By opening the HTML document in a browser and scrolling down through it, students can see how their city design and their reasoning has changed, and how it has resulted in different outcomes in their city. Two groups of students in each class were also asked, at the end of the activity, to reflect upon and present their iterative design process to their fellow students by showing them data from the HTML document. We are currently in the process of analysing the data from the students' process of building and debugging their cities (and thinking), and we will be publishing our findings in the coming years.

« 18 » **Bryan Guo** raises a question about the design of the simulation itself. He first argues that the role of stochasticity in agent-based models of complex systems makes causal relations less obvious or clear, and then asks, given the stochastic growth of the resulting city, whether students can even know that their cities are a result of *their* design decisions, and not randomness.

« 19 » When designing the algorithm that citizens in the model use to decide where to live, we validated the model by running the same city designs but with different random seeds. We wanted to see if we got any kind of equivalence between each run, as the opposite would have indicated that the results were simply random. The results of our validation efforts were, as **Guo** would probably expect, not the same every time. However, we do not think this was a problem for the purpose of our learning activity. First, while we saw *spatial* differences in the distributions of wealth, commute times, or access to parks in the model, they resulted in similar distributions as *measured by policy outcomes*. This is not to say that we achieved a numerical equivalence between runs. However, we did achieve a distributional equivalence, and we therefore concluded that students' urban planning decisions were sufficiently and systematically determining the outcomes of the cities (Wilensky & Rand 2007). Second, even if the same design had evolved into completely different outcomes, these would have still been the consequence of the underlying mechanisms – the ones that we wanted students to reflect on. As long as cities grow in a way that is explainable by the underlying mechanisms, the consequences of students' design decisions will be within reach of students.

« 20 » In his Q3, **Valente** asks about the role of the teacher during the learning activity and what kind of approach the teacher can take to intervene at the classroom level.

« 21 » We claim that teachers can and should play an active role over the course of the activity. This requires a focus on developing materials that explain the model's runtime behaviour in ways that enable teachers to spot what problems the students are facing in the construction of their cities. The classroom implementation that we reported on in the target article was the first in a series of implementations, and we have since then started to collaborate with teachers on developing teacher guides aimed at this. These materials give examples of how and why particular, common student city design patterns result in particular outcomes, and what kinds of changes are required in the design of the city in order to improve on the outcomes, and *why*. Teachers are given suggestions on how to facilitate students attending to the problems in their own design, so that they can understand what the bugs are in their students' thinking, and how to progress from there.

Use of ARM findings in practice

« 22 » In his Q2, **Valente** raises another question relating to the role of the teacher's during the activity, i.e., how teachers can use the ARM results to help students debug their misconceptions.

« 23 » Our present analysis and data do not enable us to conclude to what extent teachers can effectively help students in their debugging process. Our longer-term plan is to better understand the evolution of students' level of sophistication in reasoning about complex systems, and to potentially identify patterns in their initial, naïve understanding of a system, and the pathways through which they progress as they debug their cities, and in their thinking. Once this work has been developed more, it may be possible to use these patterns as diagnostics for determining where in their development students are, which causal relations they seem to be underemphasizing in their explanations, and what are potentially the best pathways for them to traverse in order to improve their reasoning.

References

Anderson E., Gupta A., Philip T. M., Markauskaite L., Kali Y., Goodyear P., Hjorth A. & Levrini O. (2018) Knowledge analysis outside the STEM classroom. In: Kay J. & Luckin R. (eds.) Rethinking learning in the digital age: Making the learning sciences count. 13th International Conference of the Learning Sciences (ICLS 2018), Volume 2. International Society of the Learning Sciences, London: 1219–1226. https://repository.isls.org/bitstream/1/596/1/269.pdf

diSessa A. A. (1993) Toward an epistemology of physics. Cognition and Instruction 10(2–3): 105–225.

diSessa A. A. (2002) Why "conceptual ecology" is a good idea. In: Limón M. & Mason L. (eds.) Reconsidering conceptual change: Issues in theory and practice. Kluwer, Dordrecht: 29–60.

diSessa A. A., Sherin B. & Levin M. (2016) Knowledge analysis: An introduction. In: DiSessa A. A., Levin M. & Brown N. J. S. (eds.) Knowledge and interaction: A synthetic agenda for the learning sciences. Routledge, New York NY: 30–71.

Hjorth A. & Krist C. (2016) Unpacking social factors in mechanistic reasoning (or, why a wealthy person is not exactly like a grey squirrel). In: Looi C. K., Polman J. L., Cress U. & Reimann P. (eds.) Transforming learning, empowering learners. The International Conference of the Learning Sciences (ICLS 2016), Volume 2. International Society of the Learning Sciences, Singapore: 894–897. https://repository.isls.org/bitstream/1/337/1/130.pdf

Kapon S. & diSessa A. A. (2012) Reasoning through instructional analogies. Cognition and Instruction 30(3): 261–310.

Krist C. & Rosenberg J. (2016) Finding patterns in and refining characterizations of students' epistemic cognition: A computational approach. In: Looi C. K., Polman J. L., Cress U. & Reimann P. (eds.) Transforming learning, empowering learners. The International Conference of the Learning Sciences (ICLS 2016), Volume 2. International Society of the Learning Sciences, Singapore: 1223–1224.

Levrini O. & diSessa A. S. (2008) How students learn from multiple contexts and definitions: Proper time as a coordination class. Physical Review Physics Education Research 4(1): 010107. https://link.aps.org/pdf/10.1103/PhysRevSTPER.4.010107

Philip T. M. (2011) An "ideology in pieces" approach to studying change in teachers' sensemaking about race, racism, and racial justice. Cognition and Instruction 29(3): 297–329.

Sherin B., Krakowski M. & Lee V. R. (2012) Some assembly required: How scientific explanations are constructed during clinical interviews. Journal of Research in Science Teaching 49(2): 166–198.

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

Wilensky U. & Rand W. (2007) Making models match: Replicating an agent-based model. Journal of Artificial Societies and Social Simulation 10(4): 2. http://jasss.soc.surrey.ac.uk/10/4/2.html

Received: 6 July 2019
Accepted: 7 July 2019