

Article

Features of Knowledge Building in Biology: Understanding Undergraduate Students' Ideas about Molecular Mechanisms

Katelyn Southard, Tyler Wince, Shanice Meddleton, and Molly S. Bolger

Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ 85721

Submitted May 18, 2015; Revised November 6, 2015; Accepted November 16, 2015
Monitoring Editor: Ross Nehm

Research has suggested that teaching and learning in molecular and cellular biology (MCB) is difficult. We used a new lens to understand undergraduate reasoning about molecular mechanisms: the knowledge-integration approach to conceptual change. Knowledge integration is the dynamic process by which learners acquire new ideas, develop connections between ideas, and reorganize and restructure prior knowledge. Semistructured, clinical think-aloud interviews were conducted with introductory and upper-division MCB students. Interviews included a written conceptual assessment, a concept-mapping activity, and an opportunity to explain the biomechanisms of DNA replication, transcription, and translation. Student reasoning patterns were explored through mixed-method analyses. Results suggested that students must sort mechanistic entities into appropriate mental categories that reflect the nature of MCB mechanisms and that conflation between these categories is common. We also showed how connections between molecular mechanisms and their biological roles are part of building an integrated knowledge network as students develop expertise. We observed differences in the nature of connections between ideas related to different forms of reasoning. Finally, we provide a tentative model for MCB knowledge integration and suggest its implications for undergraduate learning.

INTRODUCTION

Many national calls for instructional reform and decades of evidence-based research on student learning have convinced us that teaching and learning in the field of molecular and cellular biology (MCB) is difficult. The complexities of the mechanisms involved, the rapidly advancing knowledge base, and the specialized language create a daunting environment for learners and a difficult task for instructors. Several features may contribute to difficulties of learning in MCB, such as the “invisible” and “inaccessible” nature of the

entities studied (Gilbert *et al.*, 1982; Duncan and Reiser, 2007), the necessity of mapping across physical and ontological levels (Marbach-Ad and Stavy, 2000; Duncan, 2007; Duncan and Tseng, 2011; Van Mil *et al.*, 2013), and the interrelated and overlapping nature of complex phenomena (Duncan, 2007; Duncan and Reiser, 2007; Duncan and Tseng, 2011; Van Mil *et al.*, 2013). However, literature on the nature of knowledge and conceptual change in MCB is less well developed (with notable exceptions, including such work as Venville and Treagust, 1998; Duncan, 2007). Unanswered questions include: How do students draw connections between ideas to build expert knowledge networks? How do they organize ideas when learning about cellular phenomena? What are the underlying mechanisms of conceptual change that may lead to scientifically non-normative ideas and misconceptions? Building from cognitive learning science theories and the literature describing domain-specific reasoning strategies, we explore undergraduate students' understanding of foundational MCB principles, specifically the processes of DNA replication, transcription, and translation.

We propose that the complex nature of MCB requires learners to build highly integrated knowledge networks that are both productively organized and flexibly dynamic

CBE Life Sci Educ March 1, 2016 14:ar7

DOI:10.1187/cbe.15-05-0114

Address correspondence to: Molly S. Bolger (mbolger@email.arizona.edu).

© 2016 K. Southard *et al.* CBE—Life Sciences Education © 2016 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

to engage in problem solving and explanation building. We hypothesize that learners must integrate pieces of knowledge, manipulate connections between ideas, and build a cohesive view of the relationships between phenomena to truly build understanding of mechanisms in molecular biology. Although much work has been done to characterize the nature of knowledge among novice learners in science (Gobbo and Chi, 1986; diSessa, 1988; Vosniadou and Brewer, 1992; Southerland *et al.*, 2001; Chi, 2013) and experts in science (Chi *et al.*, 1981; Machamer *et al.*, 2000; Darden, 2002, 2008; Craver, 2002), we know much less about how scientific knowledge develops as students move through an undergraduate major.

In this study, we provide a new lens with which to explore undergraduate students' understanding of concepts within molecular biology courses. By characterizing undergraduate students' understanding of key molecular principles, using the knowledge-integration approach to conceptual change, we explore the nature of knowledge and the ways in which it is organized by students, productively and unproductively, as they begin to grapple with the complexity of molecular phenomena.

THEORETICAL FRAMEWORK

The Theory of Knowledge Integration

Learning is a dynamic process involving the acquisition of new ideas, the development of connections between ideas, and the reorganization of prior knowledge (Clark and Linn, 2003, 2013). Much research in the learning sciences has focused on conceptual change, the process by which new knowledge is built (Özdemir and Clark, 2007; Vosniadou, 2013b). The knowledge-integration theory of conceptual change focuses on the ways in which ideas are *sorted*, *connected*, and *integrated*. Knowledge integration is, by definition, the process by which learners not only add new ideas or facts but also sort through connections and move toward developing a cohesive mental model of a phenomenon (Clark and Linn, 2003, 2013). In doing so, one must *sort ideas into appropriate categories*, and *make connections between ideas*, thereby *integrating ideas to build a network of knowledge*. Knowledge integration is also significantly impacted by the amount of instructional time allocated to important topics. Clark and Linn (2003) demonstrated that significant enhancement of knowledge integration around foundational concepts in thermodynamics resulted from increased instructional time in middle school classrooms, thereby underscoring the idea that deeper learning takes time.

A key component of the theory of knowledge integration is that learners must sort through ideas to construct a cohesive view of a phenomenon. Examples of this sorting include the processes of coalescence and differentiation. Coalescence is the process by which two ideas are merged or combined, whereas differentiation involves the splitting of an idea into separate elements (Clark and Linn, 2013). For example, combining the ideas of heating and cooling into one model of thermal equilibrium would be the "coalescence" of two previously separate ideas. Alternatively, the separation of heat and energy from the idea of temperature would be "differentiation." While knowledge integration, including the active sorting and restructuring of ideas, is necessary for build-

ing scientific knowledge, not all actions of restructuring, integration, coalescence, and/or differentiation are immediately productive (Clark and Linn, 2013). That is, as new connections are made, they may reinforce misconceptions or create non-normative scientific explanations. Non-normative explanations are defined as student ideas that would not be generally accepted by the scientific community, though they may represent logical trains of thought on the part of the student. These are ideas that would typically be considered "incorrect" by an instructor. By contrast, a normative explanation reflects current norms of the scientific community and would generally be considered "correct" by an instructor.

Another key component of the theory of knowledge integration is the creation of connections between ideas. For example, when learning about "mutations," students should form connections between this idea and the concepts of "gene" and "evolution." Some connections made by students are spontaneous and temporary, whereas others are stable and persistent. Additionally, not all connections give rise to scientifically normative ideas, but they can also give rise to conflicting and dissenting ideas within a conceptual ecology (Clark and Linn, 2013).

Possibly the most complex component of the theory of knowledge integration is the integration of ideas to build complex knowledge structures. According to Clark and Linn (2013), at its most basic level, "integration involves creating or reinforcing the connection between two ideas" (p. 522). Along these same lines, when investigating students' understanding of science at different levels of knowledge integration, Clark and Linn (2003) defined a "nuanced" level of understanding in which answers "involve not only the normative ideas but also important connections to other normative ideas or evidence" (p. 459). It is also important to consider that the process of integrating knowledge, including restructuring and reorganizing ideas, leads to complex, dynamic, and emergent knowledge networks (Brown, 2013). As researchers, in order to explore how students are integrating ideas, several conceptual change theories agree that "we need to move away from thinking about conceptual changes as involving single units of knowledge to systems of knowledge that consist of complex substructures that may change gradually in different ways" (Vosniadou, 2013a, p. 12). To study these complex substructures, we must approach knowledge about this topic as a dynamic ecology of ideas, constantly under reconstruction and restructuring.

In discussing how ideas are sorted, connected, and integrated, it is important not only to look at what ideas are connected but to explore the nature of how the ideas are connected. Students' motivation for creating particular connections, the robustness of those connections, and the nature of the reasoning that forms a connection between two ideas are important features of knowledge integration that, to our knowledge, have been largely unexplored. We propose that connections between ideas in molecular biology are constrained by multilevel, molecular mechanistic reasoning that may guide developing knowledge to reflect the nature of biological phenomena.

The Nature of MCB Knowledge

In MCB, specific forms of reasoning are required to navigate the unique complexities of the domain; this process often

includes employing reasoning strategies that mirror the nature of the phenomena themselves. “Mechanistic reasoning” is one such type of reasoning required to study the interacting molecular mechanisms that underlie biological phenomena in the field of molecular biology (Machamer *et al.*, 2000; Russ *et al.*, 2008). Understanding these complex biological mechanisms involves 1) recognizing the involvement of specific entities (objects), 2) recognizing the activities created by the interactions of the entities, and 3) understanding the temporal and spatial interactions and organization of these components that create the biological phenomena (Machamer *et al.*, 2000; Russ *et al.*, 2008; Van Mil *et al.*, 2013). For example, thinking about the molecular mechanism of translation (the synthesis of proteins based on a mRNA molecule) requires the recognition of key entities such as “RNA transcript,” “ribosome,” “tRNA,” and “amino acids,” among others. It also involves recognizing activities such as “binding” of the tRNA to the RNA transcript and ribosome and “recognition” of the ribosome binding site on the RNA by the ribosome. And finally, an understanding of the dynamic sequence of events and the necessary physical organization of the molecules allows for a big-picture understanding of the phenomenon. Craver (2001) claims that understanding a mechanism involves “understanding how one activity leads to the next through the spatial layout of the components and through their participation in a stereotyped temporal pattern of activities from beginning to end” (p. 61). Explanations of molecular mechanisms involving these features can be described as containing “productive continuity,” in which entities and activities are connected without gaps from setup to termination conditions (Machamer *et al.*, 2000).

In MCB, molecular mechanisms require “multilevel reasoning.” Namely, organization at submolecular, molecular, subcellular, and cellular levels must be considered. Additionally, Hartwell *et al.* (1999) note that “a discrete biological function can only rarely be attributed to an individual molecule ... most biological functions arise from interactions among many components” (p. C47), thereby proposing “functional modules” as an organizational level between molecules and cells. Functional modules are defined as an ensemble of molecules that are considered a unit due to their cooperative function rather than a shared spatial organization (Hartwell *et al.*, 1999; Hofmann *et al.*, 2006; Van Mil *et al.*, 2013). For example, an intracellular signaling cascade or mechanisms like DNA replication and protein synthesis can be considered functional modules (Hartwell *et al.*, 1999). Though functional modules are composed of an ensemble of molecules, they must often be treated as a single entity with a collective function when placed within a biological mechanism (Hartwell *et al.*, 1999; Van Mil *et al.*, 2013).

Although mechanistic reasoning is needed to explain complex phenomena in molecular biology, even young students may use mechanistic reasoning to explain other phenomena (Metz, 1985; Lehrer and Schauble 1998; Bolger *et al.*, 2012). However, this form of reasoning is sometimes difficult to cultivate among novices (Russ *et al.*, 2008). Heuristic or rule-based reasoning for scientific explanation has been described among experts and novices (Guindon, 1990; Gigerenzer and Gaissmaier, 2011). Although heuristic reasoning can be productive and efficient, in some cases reliance on a rule without comprehending the underlying mechanism can be a hindrance to student understanding (McClary and Talanquer,

2011). If a student used a heuristic like “negative biological things are contagious,” then he or she might describe mutations as “spreading” between the cells of an organism. In contrast to mechanistic reasoning, students may also utilize teleological reasoning to explain biological phenomena; for example, explaining that a plant grows in that direction because it needs sunlight (Southerland *et al.*, 2001). Likewise, they may focus on the static or structural features of a situation instead of the dynamic or functional features when making predictions; for example, predicting the motion of a lever based on its shape rather than visualizing its motion (diSessa, 1993; Bolger *et al.*, 2012). Thus, developing knowledge in science involves learning what types of explanations are most appropriate for a given situation.

Molecular Mechanisms

We selected three molecular mechanisms to explore knowledge integration in undergraduate MCB: DNA replication, transcription, and translation. These three molecular mechanisms were selected based on their foundational importance to biological phenomena and their common coverage in undergraduate MCB courses. Owing to their centrality and foundational nature, mitosis and the central dogma (Crick, 1970) are often addressed in K–12 grades, but they are revisited and built upon in more molecular mechanistic detail during undergraduate introductory biology courses. They are then used as foundational principles with which to understand many biological phenomena in upper-division courses such as genetics, cell biology, immunology, developmental biology, and neurobiology. Despite their role as a platform for further learning in biology, the biological mechanisms by which genes are inherited and give rise to traits are often misunderstood by students (Duncan, 2007).

AIM OF THE STUDY

We carried out this study to characterize knowledge integration in undergraduate MCB at different levels of educational development. To explore this, we asked the following research questions:

Research question 1: How do undergraduate students *sort*, *connect*, and *integrate* ideas in the context of learning about transcription, translation, and DNA replication?

Research question 2: What is the *nature of connections* between ideas in undergraduates’ understanding of transcription, translation, and DNA replication?

METHODS

Study Population

We recruited students for participation in this study on a volunteer basis from one of three courses in the Molecular and Cellular Biology (MCB) program at the University of Arizona: Introductory Biology ($N = 22$), Molecular Genetics (upper-division, $N = 23$), and Cell and Developmental Biology (upper-division, $N = 8$). Most students enrolled in the introductory courses are sophomores or juniors, and most upper-division students are juniors or seniors. We selected these three courses for recruitment, because each course is required for graduation with an MCB major and each course

either explicitly covers the molecular mechanisms of DNA replication, transcription, and translation, and/or makes explicit use of these concepts. Of the 22 introductory students who volunteered for the study, 13 were science majors, two were non-science majors, and seven were MCB majors who are required to continue to the upper-division courses we sampled. Sections of the introductory course are taught by several different instructors with individual design and instructional styles. Because we were interested in exploring the ways in which individual students grapple with learned information, we recruited students from four introductory sections. Each section was taught by a different instructor who routinely used some combination of active-learning techniques such as clicker questions, think-pair-share questions, small-group work, and whole-class discussions. Each upper-division course is taught by a team of instructors, and the teams for both courses intentionally incorporate components of active-learning pedagogy into the courses, including clicker questions, think-pair-share questions, small-group work, data-interpretation studies, analysis of primary literature, and problem-solving sessions.

Discussion of the molecular mechanisms of DNA replication, transcription, and translation in the introductory sections varied slightly based on instructor. These three target molecular mechanisms were reviewed and expanded on in the Molecular Genetics course during units about heredity of information and protein synthesis, structure, and function. The three target mechanisms were then revisited intermittently in the Cell and Developmental Biology course when needed, particularly during units on early embryonic development involving gene expression, the cell cycle and downstream effects of intracellular signaling cascades. By the end of these courses, students are expected to understand the mechanisms of DNA replication, transcription, and translation in significant detail (including, e.g., replication forks, RNA primers, origins of replication, chromosome structure, histones, transcription activators and repressors, directionality of elongation, operons, sites within the ribosome, translation elongation factors, and the process of translation termination) and to apply those mechanisms to cellular and organismal phenomena. In addition to mechanistic detail, all instructors emphasized in some way the rationale behind the use of molecular mechanisms and their connection to overall biological phenomena.

We recruited students for the study through several announcements to the class. We explicitly requested student volunteers who did and did not feel that they understood the course material well to avoid a selection bias as much as possible. The data were collected over the course of four semesters, Fall 2012 to Fall 2014, and a small subset of students participated in interviews in more than one class. Think-aloud interviews were conducted at the end of each respective course in the 2–3-wk period before the final exam. Occasionally, students were interviewed several days after the final exam due to scheduling conflicts. All data collection and analysis was approved by the institutional review board at our university.

Think-Aloud Interviews

We conducted semistructured, clinical think-aloud interviews lasting approximately 1 h with each student participant and

Concept Assessment Score

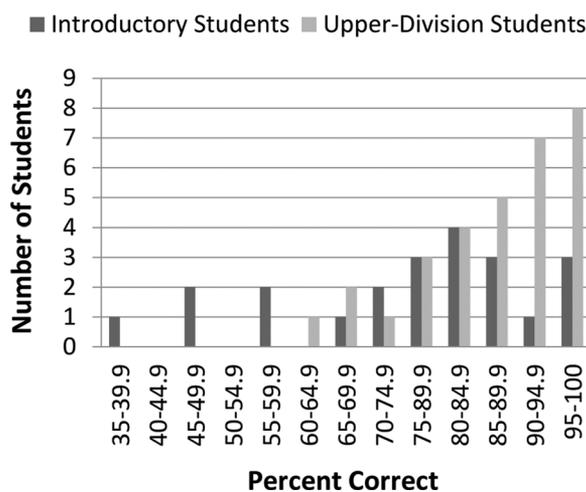


Figure 1. Upper-division students ($N = 31$) perform better than introductory students ($N = 22$) on a basic seven-question concept assessment (two-tailed independent-samples t test, $p = 0.006$).

audio-recorded the interviews. Interviews were conducted by one of six trained interviewers in the research group using a standardized interview protocol (see the Supplemental Material for the protocol). During the interviews, students started by completing a basic concept assessment designed to test conceptual knowledge in a familiar test-like format (see the Supplemental Material). The concept assessment included seven questions using a mix of multiple-choice, short-answer, and fill-in-the-blank questions. Results of the concept assessment (see Figure 1) show a spread of performance, with substantial overlap between the performance of upper-division and introductory students. Concept assessment items were assigned point values, and each students' concept assessment was graded for correctness. Students' reported concept assessment scores were calculated as the percentage of points earned out of total possible points. Concept assessment scores were on average higher ($p = 0.006$) in the upper-division group ($89 \pm 1.7\%$, $N = 31$), and variance was lower ($p = 0.015$) than in the introductory group ($76 \pm 3.7\%$, $N = 22$). Owing to the apparent skewness in the distribution of students' individual percentage concept assessment scores, as revealed by a histogram plot (Figure 1), this tendency toward higher scores in the upper division was confirmed by the Mann-Whitney U -test ($p = 0.006$). (See *Statistical Analysis* section.)

Similarly to Clark and Linn (2013), we used explanations and causal descriptions to probe for "underlying (although often unarticulated) views that shape students' thinking, explanations, and predictions" (p. 521). During the interview, we asked each student to create explanations and causal descriptions of the processes of DNA replication, transcription, and translation. Specifically, we asked students to describe what they know about the mechanism (or process) of each of the three target mechanisms, including descriptions of what is physically transpiring, the molecular "players," and where the process is happening within a cell. We did not specify whether students should describe these mechanisms

for eukaryotic or prokaryotic cells; however, most students focused on eukaryotic mechanisms, with only some students choosing to talk about prokaryotic mechanisms as well. Next, we asked students to explain why a cell undergoes that particular process. We asked students to explain their thinking by verbally describing their ideas, and students were often asked to expand their descriptions. Interviewers asked both standardized and individualized follow-up probes to clarify student thinking. We encouraged students to make drawings or sketches when they felt it would aid their verbal description and to be honest about the things they knew and did not know. The interview protocol also included a selected assortment of other tasks, including discussing mutations and problem-solving activities.

One additional task during the interview included the creation of a concept map, for which we asked students to create a free-form concept map using 20 standardized cellular and molecular terms. A brief introduction and instructions on concept mapping were given to students before they began the activity (see the Supplemental Material), and most students were familiar with concept mapping from previous courses. The concept map terms are as follows: cell, phenotype, nucleic acid, gene regulation, inheritance, enzyme, membrane, protein, DNA, transcription, ribosome, replication, mutations, gene, genetic disease, amino acid, neuron, translation, RNA, and skin cell. We asked students to verbalize their thinking as they created a concept map using as many of the given terms as possible. Students were required to label connections between terms with “linker” words, terms, or phrases to describe the relationship. More details about the selection of terms and implementation of the concept maps are provided in the Supplemental Material. After students finished creating their concept maps, we asked them to verbally guide us through the concept maps and describe the relationships they had indicated between terms. Students were also given the opportunity to add terms during this process if they felt it would enhance their concept maps, but most students chose not to add terms.

Analysis

Case Study Analysis. We performed case study analysis of the student interviews, using transcripts made from interview audio and any student artifacts (including drawings and notes about gestures) to investigate patterns of reasoning in student interview transcripts. Specifically, we closely examined eight students with diverse levels of conceptual understanding, wrote and discussed claims that might be made about their patterns of reasoning, and compared cases to highlight what seemed to be essential themes and differences between individuals. We will not describe the detailed results from these case studies here, but patterns observed in the case study analysis provided insights to aid in the setup of our coding analysis and pointed us toward portions of the interview protocol that were most useful for addressing our research questions.

Coding Analysis

Mechanism Descriptions. The structure of the interview called for students to first describe the molecular mechanistic events of DNA replication and then describe why a cell would undergo the process of DNA replication. Next, students were asked to describe the molecular mechanistic

events of transcription and explain why a cell would undergo transcription. Finally, following this same format, students were asked about translation. We coded each interview transcript for the inclusion of mechanism-appropriate versus mechanism-inappropriate entities. For example, the mechanism of DNA replication would be marked as including “mechanism-appropriate entities” if the student’s description included only entities that are agreed upon by the scientific community as participating in that mechanism, such as DNA polymerase, dNTPs, ligase, or topoisomerase. However, the mechanism was marked as including “mechanism-inappropriate entities” if the student’s description included one or more entities that were inappropriate to the mechanism of DNA replication, such as the involvement of ribosomes or flagella. To investigate the prevalence of misconnections between the three target mechanisms and their roles in the cell, we performed further coding analysis. Student responses to the question “Why does a cell undergo [target mechanism]?” were coded for scientific accuracy or plausibility (scientifically normative connections). Coding analysis was performed independently by three coders, with 78% agreement among coders (number of agreed-upon instances/total number of coded instances). Final codes were the result of consensus among coders. (See the Supplemental Material for full coding scheme.)

Concept Map Connections. Coding analysis of the nature of connections between terms was applied to all student concept maps. A connection was coded as “type 3” if it indicated or implicated a relationship between the terms based on function, mechanism, causation, or an action. A connection was coded as “type 2” if it indicated a relationship between terms based on a vague action, a structural explanation (location, composition, etc.), or a vague role in a process, or if it indicated that one term was a type or example of the other term. Finally, a connection was coded as “type 1” if it was transparently associative (e.g., “associated with”) or was left blank (line or arrow indicating connection but lacking descriptive word or phrase). Terms that were not connected by the student were not coded. Coding analysis was performed independently by three coders with an agreement of 86%. Final codes were the result of consensus among coders. (See the Supplemental Material for coding scheme and an example of a student concept map.)

Iterative Transcript Read-Throughs. We performed systematic read-throughs of all student transcripts for the portions of the interview we identified through case studies as most useful to address our research questions. Read-throughs provided further evidence of themes discovered through case study analysis and were used to triangulate qualitative patterns observed in the coding analysis. All students were given pseudonyms. Quotes from individual students in this paper are identified by pseudonym and course level.

Statistical Analysis. Concept assessment scores were recorded for each student as the percentage of points earned (based on correctness) out of total possible points. Concept map scores were recorded for each student as the average type number (type 1 = 1, type 2 = 2, and type 3 = 3) of the connections drawn by that student. Concept assessment and concept map scores were compared between introductory and upper-division groups using two-tailed,

independent-samples *t* tests. Because Levene's test indicated unequal variances between groups ($p = 0.015$ for concept assessment and $p = 0.003$ for concept maps), Welch's unequal variance version of the *t* test was used. Histogram plots of individual concept assessment scores and individual concept map scores, by group, suggested nonnormal distributions; therefore, differences between groups were also compared using the nonparametric Mann-Whitney *U*-test. Results of the independent-samples *t* test were used to describe differences between groups, as there were no instances in which the *t* tests were in conflict with the Mann-Whitney *U*-tests.

For each of the three target mechanisms (DNA replication, transcription, and translation), student descriptions were coded as including "mechanism-appropriate entities" or "mechanism-inappropriate entities." Similarly, student connections between the mechanisms and their roles in the cell were coded as "normative" or "non-normative." Proportions of descriptions including "mechanism-appropriate entities" were compared between introductory and upper-division groups using Fisher's exact test (which is more appropriate than the chi-square test for the small sample size). Similarly, proportions of "normative" connections between each mechanism and its role in the cell were compared between introductory and upper-division groups using Fisher's exact test. Statistical significance was declared when the p value was < 0.05 ; however, actual p values are reported to provide more detail. Computations were performed using IBM SPSS for Windows, version 22.

RESULTS

Introduction

We have organized our results into two parts. In part 1, we investigate various components of knowledge integration. In part 2, we examine the nature of the connections students make between ideas. Data presented include qualitative coding analysis of interview transcripts (in part 1) and students' concept maps (in part 2) and samples of student thinking during interviews that exemplify patterns we found in qualitative read-through analysis of transcripts (in parts 1 and 2).

Part 1. Knowledge Integration: Sorting through Ideas about Molecular Mechanisms in Biology

Sorting Molecular Entities into Appropriate Categories. As we previously described, a key component of the theory of knowledge integration is that learners must sort through ideas to create cohesive knowledge networks. One aspect of creating knowledge networks is the restructuring of connections between ideas, including differentiating one idea into two distinct components. Therefore, we first examined the ways in which students sort ideas about molecular entities into mental categories. To explore this, we investigated the ways in which students were discussing molecular entities in their descriptions of the three target mechanisms: DNA replication, transcription, and translation.

We found that students varied in the extent to which they included scientifically normative (mechanism-appropriate) entities when describing molecular mechanisms. For the mechanism of DNA replication, 72% of students were able to create molecular mechanistic descriptions that included

only scientifically normative, mechanism-appropriate entities; 69% of students were able to create descriptions of transcription with mechanism-appropriate entities; and 81% of students were able to create descriptions with mechanism-appropriate entities for translation. No statistically significant difference was found between introductory and upper-division students (Fisher's exact test: DNA replication, $p = 0.76$; transcription, $p = 0.77$; translation, $p = 0.27$). Figure 2 shows the percentage of students who included *only* mechanism-appropriate entities in none, one, two, or three of the three target molecular mechanisms (transcription, translation, and DNA replication). More than half of the students interviewed included mechanism-inappropriate entities in one or more of the three target mechanisms. No significant difference was found between introductory and upper-division students (Mann-Whitney *U*-test: $p = 0.53$). This was despite the fact that upper-division students had a significantly higher performance on a written concept assessment (see *Methods*).

When students included a mechanism-inappropriate entity in their descriptions of one of the three mechanisms, 90% were instances in which students included an entity from one of the other two mechanisms (e.g., discussing the role of the ribosome, which is a translation-specific molecular complex, in the process of DNA replication). Coders were careful not to code "mis-labeled entities" as inappropriate entities. Examples of this include vague descriptions (e.g., "that thing that comes in and adds the nucleotides"), when students corrected themselves (e.g., "DNA polymerase, I mean RNA polymerase"), or when a student obviously misspoke

Sorting Entities into Appropriate Mental Categories

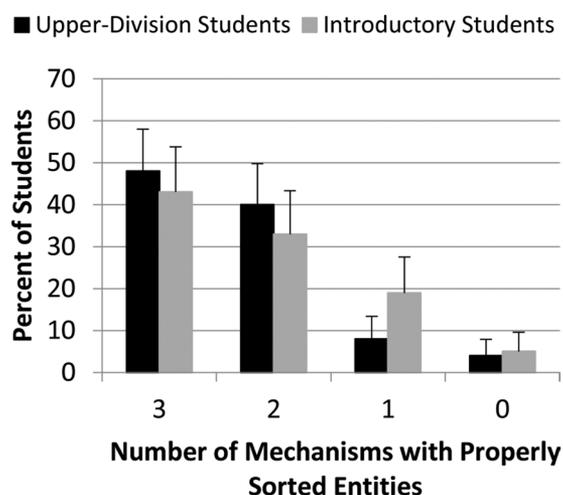


Figure 2. Percentage of students within the group (introductory students: $n = 21$; upper-division students: $n = 25$) who created molecular mechanistic descriptions of none, one, two, or three of the target mechanisms (DNA replication, transcription, and translation) using mechanism-appropriate entities. Error bars represent the calculated SE. (Seven students who missed one of the three target mechanisms due to time constraints in their interviews were not included.)

(e.g., switching the words “transcription” and “translation” but describing the intact mechanisms).

These coding results suggest that some students have difficulty sorting molecular entities into appropriate mental categories that reflect the nature of these mechanisms. While the results of this coding analysis showed us whether entities were being appropriately sorted, qualitative analysis of transcripts during read-throughs revealed how the students were using the entities in their mechanistic descriptions and how they were thinking about their mechanisms. In some cases in which students missorted entities, there was evidence that students were conflating mental categories by blurring the boundaries that distinguish the three target mechanisms. This conflation of mental categories resulted in the inappropriate fusion of two or more of the target mechanisms. For example, when one student was asked to describe the molecular mechanism of DNA replication, he responded:

Line 1: “Okay so, from what I know, based off ... the cell ... so you have the mRNA that left the nucleus and I am sure it gets transcribed by the RNA and as RNA ... no as the ribosomes transcribe the RNA ... I want to say that it, like, it adds, like, I want to say it adds, like ... [trails off into inaudible mumbling].”

Line 2: Interviewer: “What are you thinking about?”

Line 3: “Now I am trying to tie the replication fork in now.”

Line 4: “So ... you have ... I want to say that the ribosome ... like ... adds ... I want to say, like, temporary base pairs to it?”

Line 5: “I don’t really know what happens there ... but all I know is ... so, like, you have this when it, like, splits ... I know, like, an animal and, like, eukaryotic ... they have, like ... let’s see ... so the DNA ... um ... so this could be, like, the leading strand and lagging ... and then there is something here ...”

Line 6: “Something that goes here ... I think it’s, um, no RN ... ‘cause all I know in the leading strand there is, like ... ‘cause, like, I don’t know which enzyme or protein ... or whatever it is that ... so, like ... I think it’s in the lagging strand and something goes along and, like, fills in, like ... it adds, like, temporary ... I don’t know what it’s called ... RNA-tides?” (Jasper, introductory student)

This student’s description indicates that he is aware that the task at hand requires him to create a molecular mechanistic description. He begins his description of the mechanism by using both biological entities and activities. His description includes several entities that are appropriate to the mechanism of DNA replication (replication fork, DNA, leading and lagging strands: lines 3–5). However, he also includes entities and activities like mRNA, ribosomes, mRNA transport out of the nucleus, and transcription, which are all features of the process of protein synthesis (involving the mechanisms of transcription and translation: line 1). The inclusion of both mechanism-appropriate entities and entities belonging to the mechanisms of transcription and translation reveals a conflation, or lack of appropriate differentia-

tion, between the functional modules of DNA replication, transcription, and translation. This phenomenon of conflation between mental categories was observed in several students’ descriptions of the target mechanisms.

In contrast, when students included only mechanism-appropriate entities in their mechanistic descriptions, mental categories appeared to be clearly defined. For example, another student’s description of DNA replication contains no evidence of inappropriate conflation with transcription or translation:

Line 1: “DNA polymerase would come in, but it can’t start by itself so it needs RNA primase to make the primers, to make those three primers in the 5-prime to 3-prime direction, and so when it does that then DNA polymerase just goes in the 5-prime to 3-prime direction, as it is going [the] helicase unwinds the DNA and it separates it.”

Line 2: “And in that direction it just goes fine, just, like, all the way. And then in the 3-prime to 5-prime direction it’s harder because you ... it’s like ... so RNA primase comes in every so often and adds in primers then DNA polymerase, because it can’t go 3-prime to 5-prime, it kinda goes in the other way and does little chunks at a time” [...]

Line 3: “And so it does those in little chunks and those fragments are called Okazaki fragments.” (Bobby, upper-division student)

In this excerpt, the student is, like the previous student, including mechanism-appropriate entities (DNA polymerase, primers, helicase, Okazaki fragments), but unlike the previous student, includes no mechanism-inappropriate entities. Thus, correct sorting of entities seemed to be associated with clear boundaries between mechanisms, as opposed to conflation of mechanisms.

Connecting Biological Ideas. Another key component of the theory of knowledge integration is the creation of connections between ideas, particularly connections that integrate knowledge elements. Therefore, we next explored the connections students made between the mechanical details of a molecular mechanism and that mechanism’s overall function. After students were asked to create a molecular mechanistic description of the three target mechanisms, we asked them to explain why a cell would undergo that particular process.

Analysis of students’ responses revealed a range of connections. Most students were able to create scientifically normative or plausible connections between the target molecular mechanisms and their roles in the cell. Seventy-four percent of students were able to create a scientifically normative connection between the molecular mechanism of DNA replication and its role in the cell; 83 and 88% of students were able to create a normative connection between transcription and translation and their roles in the cell, respectively. No statistically significant difference was found between introductory and upper-division students (Fisher’s exact test: replication, $p = 1.0$; transcription, $p = 0.13$; translation, $p = 0.39$). However, a subset of students created alternative connections between these processes and their roles in cellular function, particularly for DNA replication.

Table 1. DNA replication: connection between the molecular mechanism and its role in the cell

	Normative connection to role in the cell ^a	Non-normative connection to role in the cell ^a
Description included mechanism-appropriate entities only	29	9
Description included some mechanism-inappropriate entities	10	5

^aNumber of students. $N = 53$, combined introductory and upper-division students.

Although one might assume that students who were inappropriately sorting entities might also be creating non-normative connections to the mechanism's role in the cell, our coding analysis reveals that this is not the case. In fact, several students who included mechanism-inappropriate entities in their mechanistic description of DNA replication went on to describe normative connections to the role of that mechanism in the cell. Conversely, several students who created normative mechanistic descriptions of the process of DNA replication described non-normative connections for the role of this process in the cell. For example, John provided a normative mechanistic description that included details such as "RNA primers that allow DNA polymerase to begin polymerizing nucleotides that creates a lagging strand." However, when asked "Why does a cell undergo this process?," he replied:

"Um, to, I guess to make a copy of itself. Or, to do certain duties for example, um, protein to make proteins. Without DNA, you wouldn't be able to make proteins and without replication you wouldn't be able to make more DNA in order to make more protein." (John, introductory student)

Thus, despite the level of mechanistic understanding for DNA replication, this student connected the mechanism to protein synthesis rather than mitosis or meiosis. Coding analysis (see Table 1) showed the number of students with normative or non-normative connections between the mechanism of DNA replication and its role in the cell. Results are broken down by whether the student included entirely mechanism-appropriate entities into his or her description of the mechanism or whether mechanism-inappropriate entities were included. Fisher's exact test showed no significant association between the ability to make normative connections to the function of the mechanism and the ability to

create mechanistic descriptions with entirely mechanism-appropriate entities ($p = 0.50$).

Next, we provide some examples to further illustrate the types of connections students made between mechanisms and their biological role. When Veronica, an introductory student, was asked why a cell would undergo DNA replication, she answered, "Well, it's going to have to go through mitosis and like split up all of its organelles and everything, so it's going to need to have a copy of DNA. So it just like makes a replication [copy] of itself, and so it'll split it in half like two equal [copies]." Here, this student is creating a normative connection between her previously described ideas about the mechanism of DNA replication and the overarching phenomenon of cellular division. Several students created similarly normative connections between the process of translation and protein synthesis, as demonstrated here with another student's description of the role of translation; Nikki, an upper-division student, says, "To make a functional, to make a protein out of the mRNA so that it [the protein] can function in the cell."

Although such connections are basic components of these biological phenomena, a significant subset of students described alternative connections (see Table 2). For example, when explaining why a cell undergoes DNA replication, Lydia says:

"That is a good question ... um, probably to keep the DNA, that is kind of going around in the nucleus, new and have it be able to be used by all these other things that DNA is used for. So there is lots of DNA that can be used for replication, lots of DNA for transcription, lots of DNA to be used for mitosis I guess, too." (Lydia, introductory student)

In this example, Lydia implies that DNA replication is needed to make sure the cell has "enough" DNA for all the

Table 2. Non-normative connections between DNA replication and its cellular role: what are they saying?

Heuristic	Students ^a	Examples of heuristic ^b
DNA needs to be kept "fresh," "protected," "young"	6/14	DNA disappears (Rachel, UD) DNA wears out (Brittany, UD) Risky to use original DNA, so make copies (Amy, UD)
Need "enough" DNA to make proteins or use for other processes.	8/14	Needs lots of DNA for replication, transcription, translation, and mitosis (Lydia, IN) Need extra copies to make proteins (Maria, UD; Amy, UD; John, IN)
The cellular "purpose" of the process of DNA replication is to create proteins	2/14	Purpose of DNA replication is to make proteins (John, IN; Jasper, IN)

^aMost students displayed more than one heuristic in their non-normative explanations.

^bUD, upper-division students; IN, introductory students.

cellular processes that include the DNA molecule. In doing so, she seems to form an inappropriate causal connection between DNA replication and transcription. Like this student, several other students across the study population inappropriately connected the mechanism of DNA replication to protein synthesis in a variety of ways. In fact, Maria, an upper-division student who successfully completed upper-division classes in the major, answered “Um, why? It just does. That’s, um ... well it, like, DNA needs several copies I guess to ... to develop, like ... develop, to develop, like, those amino acids, and to form those um ... proteins.” We suspect that her explanation also represents a non-normative connection between the mechanism of DNA replication, involved in cellular division, and the phenomenon of protein synthesis. This inappropriate connection between DNA replication and protein synthesis lends further support to the idea that some students conflate these mechanisms.

Further investigation of the student answers that displayed non-normative connections for DNA replication revealed that these students often use heuristic reasoning to explain the role of this mechanism. We found four main types of non-normative connections between the mechanism of DNA replication and its role in the cell (described in Table 2). Several of the student descriptions included ideas of more than one type. These misconceptions appear to be driven by heuristic reasoning (e.g., DNA must be kept fresh, new, or protected). These heuristics can be thought of as ways of thinking that may be productive in some situations but serve to reinforce non-normative connections in this case. For example, the idea that cells or molecules age over time and need to be replenished is scientifically normative, but this same idea might lead some students to think that DNA replication is an ongoing process of renewal instead of a controlled process that occurs at a distinct time in the life of a cell to provide identical copies of DNA to two newly replicated cells. These examples provide evidence that, as students build knowledge of biological mechanisms, they may form connections between ideas that are not scientifically correct but yet are based on forms of reasoning that make sense to the student at that time (as opposed to a lack of reasoning).

Integrating Ideas to Build a Network of Biological Ideas. Finally, we explored the integration of biological ideas by analyzing the ways in which students were connecting the molecular mechanisms of DNA replication, transcription, and translation to other biological ideas. By analyzing student descriptions of the three target molecular mechanisms, we discovered that a few students immediately and reflexively made connections between the mechanisms and other relevant biological phenomena. For example, after describing a scientifically normative mechanism for DNA replication, the following student answered the question “Why does a cell undergo DNA replication?”:

“Well, it undergoes replication ... so the cell will be told when to replicate based on chemical signals, so like, if, I ... you know ... have ... if my skin is dying or I have a cut or something, then the cell is going to be like, we need skin cells over here, so it will replicate to fulfill some kind of task that the body needs, for example. But the replication is a controlled process, it doesn’t just replicate whenever or else it would be cancerous.

So there is a variety of signals and all that that go into replication. And basically, if the body has a need for it, then it will replicate. And cells have a life span too, and if the cell is old then it needs to be replaced, and cells will signal to each other when we need more cells or whatever.” (James, upper-division student)

Here, the student begins by making a connection to chemical signaling, which is a mechanism for cell communication and is the starting condition for many cellular responses like DNA replication. He then creates connections to two related mechanisms, cellular regeneration and wound healing in the context of skin cells. Drawing these two mechanisms into a biologically plausible context (replacing dead skin cells or wound healing), he problematizes the cellular need for cellular replication, and therefore DNA replication. He then creates a connection to regulation and control by explaining the relationship between controlled cellular growth (in wound healing and cellular regeneration) and uncontrolled cellular replication (in the context of cancer). We observe this student extending beyond merely creating a normative connection or two between the molecular mechanism of DNA replication and its role in the cell. Instead, we observe this student’s mental category of DNA replication as being nested within the related biological phenomena of chemical signaling, cellular regeneration and repair, controlled cellular division, and cancer.

Similarly, after describing the molecular mechanism of transcription and the role of RNA as an information messenger for protein synthesis, another student creates a spontaneous connection to regulatory noncoding RNA:

“And then sometimes those intron[s] and, you know, other regions of the DNA in the pre-mRNA after they get spliced out, sometimes they do interesting things too, because those can become, like, micro-RNAs and things like that.” (Jenna, upper-division student)

By including these additional connections to their explanations, students were extending beyond merely incorporating more information into their mechanistic descriptions and instead were making connections they believed to be functionally or causally integrated to their explanations. For example, Sally crafted a description of her understanding of “mutations” when asked to elaborate during the concept-mapping portion of the interview. After connecting mutations to their potential origins during DNA replication, she went on to connect the idea of sequence changes to the role of mutations in creating genetic differences within a population:

“A mutation is when there is some kind of change in the DNA. And it comes about, I think, usually during replication if there’s an error made by the proteins that are performing the replication. And there’s, like, the proteins that check for errors, there’s ... they occasionally mess up. So that’s how mutations come about. They’re just, they change the structure of the protein, sometimes [...] And that can usually be bad or neutral for the cell, but sometimes it does good things, which is why we have diversity I guess.” (Sally, upper-division student)

She continues to expand on this idea by connecting it to two relevant biological scenarios in which this principle

holds true; first connecting to malaria resistance in individuals with sickle-cell anemia and then to receptors recognized by HIV:

“In northern Europe there’s, like, this entire deletion of like 32 nucleotides in one gene that codes for a protein that helps the HIV virus recognize cells. So if you get rid of that protein [by way of a mutation], it’s like a receptor protein, by deleting those nucleotides, those people actually become resistant to HIV because HIV can’t bind to the cells. So that’s pretty cool. That’s a good mutation. But usually they do things like they delete proteins, or they change them so that they are not functional, and usually that’s bad.” (Sally, upper-division student)

Through her description, we observe how Sally integrates the ideas of “DNA replication,” “proteins,” “mutations,” and “genetic differences,” and the implications of detrimental versus beneficial mutations that are solidified using two biologically relevant contextual examples. Nuanced connections like those exemplified here by James, Jenna, and Sally suggest that students may begin to build a network of biological ideas at the undergraduate level, integrating their ideas about the target mechanisms with functionally related biological phenomena, principles, and contexts.

Summary of Part 1 Results. In summary, we observed that students must develop mental categories with which to appropriately sort the molecular entities involved in each of the target molecular mechanisms. Some students appear to conflate one or more of these mental categories, leading to conflated mental models of the molecular mechanisms. Examination of how students connect molecular mechanisms to their overarching roles in the cell suggested that these two aspects of understanding molecular biology are not always linked by students. Finally, we found evidence of some students creating complex knowledge networks that reflect nested and overlapping ideas, suggesting a more nuanced understanding of these topics.

Part 2. The Nature of Connections

Through analysis of the previously described results, we uncovered several important differences in the ways in which students created connections. These differences suggested inherent differences in the nature of connections between ideas in MCB. To understand the nature of connections between ideas in undergraduate molecular biology, we took two approaches. First, we examined how students were connecting terms during a concept-mapping task. By analyzing the types of linking words students used to describe the relationship between two concept map terms, we were able to determine differences in types of connections. Second, we qualitatively examined students’ mechanistic explanations of the target mechanisms (DNA replication, transcription, and translation), with the lens of domain-specific reasoning strategies defined in existing literature. Findings included several characteristics of the connections between ideas *within* the mechanistic descriptions created by the students about DNA replication, transcription, and translation.

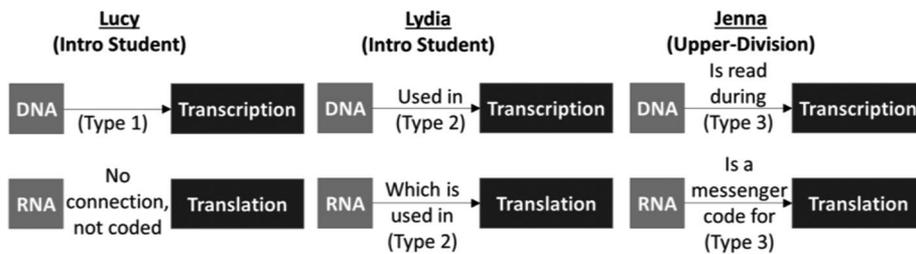
Characterizing the Nature of Connections through Concept Mapping: Associative, Functional, and Causal Connections. We examined the concept maps created by students during the concept-mapping task of the interviews. This revealed differences in the types of connections drawn between the same ideas (provided terms) by different students. A coding scheme was developed and applied to all student concept maps to systematically look at the nature of connections created between terms (Figure 3, A and B). Connections were coded as type 3 connections when linker words indicated functional, causal, mechanistic, or action-based ideas. Linker words that indicated vague actions or roles in a process, a type or example, or a structural indication (location, composition, etc.) were coded as type 2 connections. And finally, associative or blank connections were coded as type 1.

Results of the coding analysis indicated a significant difference between introductory and upper-division students’ average connection scores (independent-samples *t* test, $p = 0.013$). Although roughly half of connection linker terms on introductory students’ concept maps indicated associative connections, only 23% of upper-division concept map connections were classified as associative (Figure 3B). The percentage of type 2 and 3 connections was higher for upper-division students compared with introductory students, 19% higher for type 2 and 8% higher for type 3. It is important to mention that these categories do not reflect the “correctness” of the connections but differences in the reasoning behind why two ideas might be related. It was possible for type 1 and 2 connections to be scientifically normative and for type 3 connections to reflect misconceptions.

After establishing differences in the nature of connections between ideas through the concept-mapping activity, we returned to the students’ descriptions of the molecular mechanisms of DNA replication, transcription, and translation. By looking more deeply into the nature of connections between ideas, we found evidence to support the differences between causal, functional, and mechanistic connections in comparison with associative connections in students’ descriptions of the target mechanisms. Consider the following response when Mandy, an upper-division student, was asked why she included a particular piece of information (the directionality of synthesis of the newly made DNA strand) when describing the mechanism of DNA replication: “Um, I don’t know if I know the answer. I feel like that was just, I don’t know, I’ve just been told that ever since like eighth grade or something, so it’s like, ‘ok, 5-prime, 3-prime.’” After including the same piece of information about directionality in his explanation, Bobby was asked the same question and responded by saying “because on the, uh, 3-prime end there’s a hydroxyl group and that’s where, that’s where the phosphate of the 5-prime end of the next nucleotide, that’s where it attached to. On that 3-prime hydroxyl.” Whereas the first student’s connection between the events of DNA replication and this idea of directionality appears to be based on association, the second student’s same connection appears to be based on functional and mechanistic principles of the molecules involved. Thus, the different forms of reasoning that can be inferred from what students wrote on their concept maps are also evident in their verbal descriptions and explanations.

The results of the coding analysis showed that individual students almost always included connections in more than one category. Furthermore, different students frequently

A. Examples of Concept Map Connection Categorization



B. Concept Map Connection Categorization

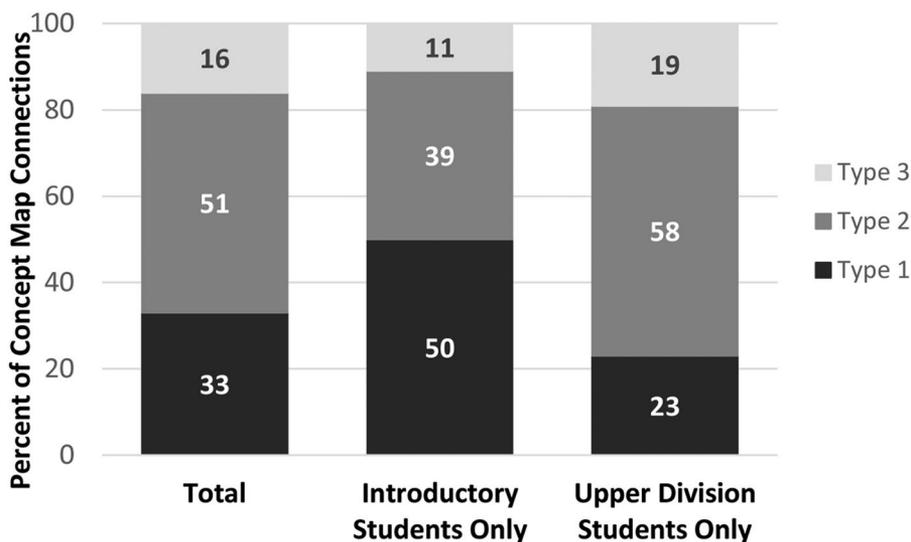


Figure 3. (A) Examples of coded connections from the concept maps of three students. (B) Percent of concept map connections made by students that were coded as functional/mechanistic (type 3); vague, structural, or categorizing (type 2); or associative or blank (type 1).

linked the same two terms with different types of connections. Thus, we hypothesize that the individual knowledge-integration process includes linking of ideas through a variety of reasoning processes, which is likely highly context-dependent and unique to the individual. However, while it may not be essential for all ideas in a knowledge network to be connected by mechanistic reasoning, our data suggest that associative connections are not typically as robust as causal, functional, and mechanistic connections.

Characterizing Connections through Molecular Descriptions: Mechanistic Reasoning. Finally, we provide our findings on the ways that students connected ideas within their descriptions of DNA replication, transcription and translation. Some students created causally connected descriptions, in which entities and activities were used to create a cohesive explanation. A small subset of students not only created productively continuous, cohesive descriptions, but also extended beyond basic causal chaining of ideas by inserting causal additions that reinforced causal connections. However, other students created fragmented explanations that were discontinuous and noncohesive, often including non-normative ideas.

Causal Connections: Mechanistic Reasoning. Machamer, Darden, and Craver propose that mechanisms that explain phenomena are produced by tracing the continuous sequence

of events, from setup to termination conditions (Machamer *et al.*, 2000; Russ *et al.*, 2008). Student descriptions that were considered causally connected recounted a sequence of molecular events in which activities are carried out by entities with specific properties. These causally connected descriptions were defined by a continuous sequence of events that often included the temporal and spatial organization of entities and activities, often using activities as a causal driver for the addition of the next piece of the explanation. For example consider this student’s description of the mechanism of DNA replication:

“And so when DNA is to be replicated, it’s double stranded and in order to replicate it the two strands have to be pulled apart from each other forming single stranded DNA. But still connected together. An enzyme called helicase is responsible ... I could have drawn that ... but it, [it] opens up the DNA, it forms the replication fork and then there’s various other proteins [single-stranded binding proteins] that bind to the DNA to prevent it from closing back in on its self, so you want to keep that replication bubble open. I forget what those proteins are called, but they just sit near helicase, I know, just to keep it from closing back up because of the complementary nature of the bases it’s always wanting to close so you need something to prop it open basically.” (James, upper-division student)

This student creates a description of this mechanism by recounting a sequence of molecular events in which activities are carried out by entities with specific properties. Scientifically normative physical entities associated with the mechanism of DNA replication (helicase, replication fork, etc.) are connected through a causal sequence of events. This form of reasoning has been described in the literature as reasoning in which a scientist or a science student uses knowledge about the inherent causal nature of the mechanism to articulate connections between ideas and create a mechanistic description, causally “chaining” through a sequence of ideas (Russ *et al.*, 2008). We observed that some students, like this student, in addition to causally chaining through molecular events, inserted causal additions to build more nuanced explanations (illustrated in the example by underlining). These causal additions appear to reinforce and strengthen causal connections within a chain of molecular events, giving a causal rationale or further explaining the function of a particular connection. It should be noted that explanations including causal additions to a productively continuous causal chain of molecular events included only rare instances of incorporation of mechanism-inappropriate entities. In other words, this more expert form of reasoning did not seem to be associated with mechanism conflation.

Fragmentation: Associative Reasoning. Fragmented descriptions, by contrast, were characterized by frequent disruptions to a chain of molecular events. Consider this excerpt from Jasper’s transcript:

“I don’t really know what happens there ... but all I know is ... so like, you have this when it, like, splits ... I know, like, an animal and, like, eukaryotic ... they have, like ... let’s see ... so the DNA ... um ... so this could be, like, the leading strand and lagging [drawing] ... and then there is something here ... something that goes here ... I think it’s, um, no RN ... ‘cause all I know in the leading strand there is, like ... ‘cause, like, I don’t know which enzyme or protein ... or whatever it is that ... so like ... I think it’s in the lagging strand and something goes along and, like, fills in, like ... it adds, like, temporary ... I don’t know what it’s called ... RNA-tides? I don’t know, but I have the image in my head, ... but, like ...” (Jasper, introductory student)

We describe this explanation as fragmented due to its lack of productive continuity. However, this student does appear to be aware that the task of creating a description of a biological mechanism involves creating a sequenced description of molecular events but is unable to do so. When at a loss for a productive causal connection, this student relies on associative connections to bring in related idea fragments. This connection-building strategy led to frequent disruptions in his chain of events and resulted in a description consisting of weakly associated idea fragments.

Other student transcripts revealed similarly fragmented descriptions of molecular events, in which students used associative connection making strategies. For example, when a different student was asked to describe the same mechanism of DNA replication she said:

“Replication ... gosh. It is in the, it’s starts in the nucleus. And then I want to say that it ends in the cytoplasm,

and some things that are involved are DNA of course, ribosome, like the larger and the smaller subunit, DNA polymerase, primers, which are RNA primers, helicase, which is another enzyme, and then ... so when, it’s when they are separated by ... the DNA strand.” (Emma, upper-division student)

Although her description begins with the start and termination conditions of the mechanism, she quickly reverts to listing entities associated with this mechanism. Her list includes both mechanism-appropriate and mechanism-inappropriate entities. As with the first fragmented quote, the majority of this student’s connections appear to be associative in nature. Analysis revealed that these fragmented explanations were closely associated with the inclusion of mechanism-inappropriate entities.

Summary of Part 2 Results. In summary, we observed differences in the ways in which students create connections during the concept-mapping portion of the interview. We further investigated these differences through in-depth analysis of student descriptions of the target molecular mechanisms. Our findings suggest that more robust descriptions were associated with use of mechanistic connections instead of associative connections. In other words, differences in the ways student understand molecular mechanisms can be attributed to differences in the nature of the connections they form between the components of that mechanism. Further, we posit that associative rather than mechanistic connections could serve as a barrier to a deeper understanding of these biological processes.

DISCUSSION

In this paper, we provide the first evidence for how the knowledge-integration theory can be applied in undergraduate biology. We have identified key aspects of how students in molecular biology sort, connect, and integrate ideas into a knowledge network. To our knowledge, this study represents the first effort to characterize domain-specific knowledge integration at the undergraduate level. We discovered that students must develop mental categories into which they can appropriately sort the molecular entities involved in the biological processes of DNA replication, transcription, and translation. We characterized the different ways students use associative, heuristic, and mechanistic reasoning when connecting ideas, and we observed that these differences can have consequences for the scientific appropriateness of students’ ideas. We showed how students create nuanced knowledge networks that reflect the multilevel nature of phenomena in molecular biology. Overall, our results support prior work suggesting that knowledge structures are dynamic, emergent, and complex (Brown, 2013; Clark and Linn, 2013). Importantly, our data underscore that knowledge integration is an ongoing process, through which a network of connections are built and refined. Our analysis has revealed that a snapshot of student understanding for a given topic at any point may reveal both scientifically normative and non-normative ideas. Next, we provide our model of knowledge integration, highlighting several of our main findings (see Figure 4).

I
N
T
E
G
R
A
T
I
O
N

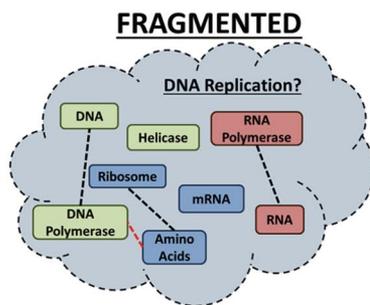
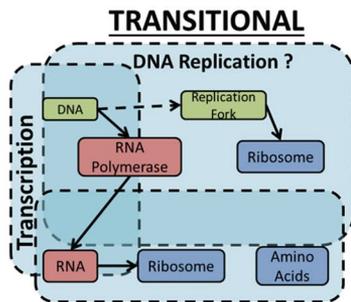
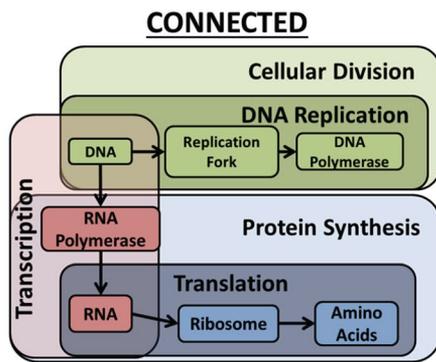
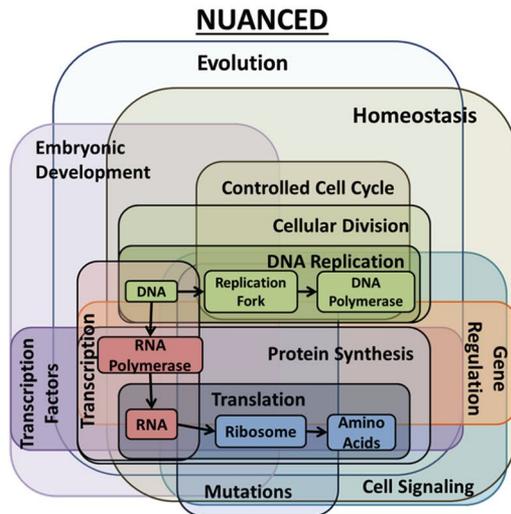


Figure 4. Theoretical model of knowledge integration in undergraduate molecular and cellular biology.

F
R
A
G
M
E
N
T
A
T
I
O
N

Model Based on Study Findings

Using the context of the molecular mechanisms of DNA replication, transcription, and translation, our model proposes several features of the process of knowledge integration (i.e., *sorting*, *connecting*, and *integrating* ideas into a knowledge network) in the domain of MCB.

First, we illustrate that fragmentation of ideas decreases as knowledge becomes more integrated. Second, the process of knowledge integration in this domain includes the creation of boundaries between mental categories, depicted as large dotted lines transitioning to solid lines that separate groups of ideas. Third, while overlapping biological ideas may initially result in conflation of mechanisms (“fragmented” section), ideas ultimately must be functionally overlapped in scientifically normative ways for students to create an integrated knowledge network (“nuanced” section). To illustrate our ideas, we describe each of the depicted levels, providing hypothetical examples of student thinking based on our study findings.

In the first “fragmented” section you can see that a hypothetical student, when asked to describe the biological processes of DNA replication, transcription, and translation, names several molecular entities (depicted by the shaded squares). Though mechanistic entities are present, the hypothetical student is unclear which entities belong with which mechanisms. In other words, the student is uncertain where the boundaries lie between the three mechanisms—where one begins and the other ends. For example, while it might be appropriate to include the molecular entity of “DNA” in both DNA replication and transcription, it is not scientifically appropriate to include the molecular entity “ribosome” when describing the mechanics of DNA replication or transcription. We propose that students with “fragmented” descriptions are likely to have undifferentiated mental categories for the three target mechanisms, leading to category conflation or inappropriate idea coalescence. Students with “fragmented” descriptions often struggled to create productively continuous explanations. These student descriptions often have frequent disruptions in their descriptions of the chain of molecular events and often name molecular entities that they associate with the mechanism while being unsure of their specific roles. Without understanding of a chain of molecular events, this student relies on vague and associative connections to attempt to build a mechanistic explanation (depicted by small dotted connections between entities).

In the “transitional” section, we illustrate the beginnings of boundary definition between mental categories for the three target mechanisms for a hypothetical student. At this level, boundaries between mechanisms are still tenuous (dotted lines), and mechanisms may or may not be connected to appropriate biological contexts (labeled “transcription,” tentative label for “DNA replication,” and missing label for “translation”). Because boundaries between mechanisms are still ill-defined, occasional errors in the inclusion of mechanism-inappropriate entities persist. For example, this hypothetical student has discussed the role of RNA polymerase both in DNA replication and transcription, when DNA polymerase is the mechanism-appropriate entity

for the activity of DNA polymerization. While vague and associative connections are still present, “transitional” descriptions include causal or functional connections between molecular events, depicted as solid arrows.

In the third section, student descriptions can be described generally as “connected.” A hypothetical student has appropriately sorted all entities into the appropriate categories for DNA replication, transcription, and translation, and these ideas have been connected to (or nested within) scientifically normative roles in the cell (protein synthesis and cellular division). Molecular entities are connected by corresponding activities that are temporally and spatially defined within the mechanistic chain of molecular events (solid arrows).

And finally, in the fourth section, we illustrate a hypothetical student’s “nuanced” network of biological ideas. Here, the student has included mechanism-appropriate molecular entities into her or his description of the three target mechanisms, defined appropriate yet flexible boundaries between mental categories, and described a productively continuous chain of molecular events in which entities have corresponding temporal and spatial activities. In addition, this hypothetical student adds functional or causal connections to relevant biological phenomena and integrates several ideas to describe a nuanced and overarching biological principle.

Our data suggest that conflation of molecular mechanisms is not uncommon among undergraduate students. This is perhaps not surprising, given the complex, multilevel nature of MCB. In particular, our model points to the importance of functional modules as a level of organization. Although functional modules have been previously described as inherent features of biological systems (Hartwell *et al.*, 1999; Hofmann *et al.*, 2006), we propose that they also serve as mental organizers. Students must integrate knowledge about the mechanics of a molecular mechanism with its overarching role in the cell or organism, thereby housing the features and mechanics of a molecular mechanism within an overarching biological function. We hypothesize that experts within the field of MCB use functional modules as mental organizers, which works to drive the appropriate categorization of new knowledge into appropriate mental categories. As our model depicts, we suggest that, for knowledge to become more integrated, students must use functional and causal drivers as organizing pillars to create an integrated network of biological ideas. Overall, the process of mental category development appears to be a complex and sensitive process. This category development requires that category boundaries be well enough defined to differentiate ideas but flexible enough to allow dynamic relationships between these ideas in order to construct a network of biological ideas.

Organizing Ideas in Molecular and Cellular Biology

Within the field of molecular biology, a biological mechanism and its role in the cell are inseparably linked. Our results suggest that, in an educational setting, these connections are not always apparent or intuitive to students. Whereas experts in the field of MCB might find it difficult to separate molecular mechanisms from their functions, our data suggest that the “mechanism” and “function” aspects of functional modules can be separated for students learning about molecular mechanisms. Our results suggest that students’ mental categories for molecular mechanisms

do not necessarily develop in parallel to the development of their understanding of the functions of the mechanisms. Although many students could describe molecular events mechanistically, in some cases, these descriptions seemed to be disconnected, that is, mechanisms without a cause.

A notable number of students struggled to connect the mechanism of DNA replication to its role in the cell. On the basis of prior work suggesting that it can be difficult for students to make connections between physical and ontological levels (Marbach-Ad and Stavy, 2000; Duncan, 2007; Duncan and Tseng, 2011), and more specifically between the molecular and cellular levels (Van Mil *et al.*, 2013), we suggest that it may be difficult to build connections between the molecular level (mechanism of DNA replication) and cellular biophysical level (process of mitosis and meiosis). By contrast, the molecular mechanisms of transcription and translation may be more easily connected to gene regulation and protein synthesis, which are also at the molecular level.

While our results indicated disconnects like these in some students’ understanding, we also explored the more nuanced connections that students made. The “core features” previously described in the literature as underlying organizers for expert knowledge structure (Chi *et al.*, 1981; Chi and Koeske, 1983; National Research Council [NRC], 2000) parallel the abstract, function-based principles we observed being used by a few students to nest and overlap concepts to create complex causal explanations. The mental categories that we describe bear similarity to what the expertise literature has defined as “chunks.” These “chunks” are collections of ideas organized based on underlying features and allow for rapid retrieval of information and enhanced problem-solving abilities (De Groot, 1965; Chase and Simon, 1973; NRC, 2000). Experts must organize ideas about molecular mechanisms by underlying functional features, “chunking” ideas and creating complex networks between chunks. Thus, students of molecular biology must build a functionally connected knowledge network that includes organization into functional modules. Further, though it has been hypothesized that having an integrated knowledge network may contribute to the ability that experts have to solve novel problems (Linn, 1995; Jonassen, 2000), this idea has not been fully explored empirically. So, while we theorize that students in our study with a more nuanced network of ideas might be better prepared to approach scientific problems, it is not within the scope of this study to claim that integrated knowledge will improve transfer.

Developing Integrated Knowledge

Our results suggest that learning is occurring between introductory and upper-division courses, as demonstrated by a significant difference in performance on a basic concept assessment. Our results also demonstrate a difference in the type of connections, on average, between introductory and upper-division students. These results suggest that more novice students in introductory courses may be relying on associative connections between ideas and that, by upper-division courses, students are beginning to move toward more functional, causal, and mechanistic connections by reconnecting or modifying existing connections. Further, the explanations provided by students in our study, at both introductory and upper-division levels, primarily included

mechanistic reasoning. Though heuristic (rule-based) explanations were occasionally used, we saw little evidence of teleological explanations. Overall, our results suggest that explanations in our study population were quite different from younger students in prior studies (Southerland *et al.*, 2001; Duncan and Reiser, 2007; Bolger *et al.*, 2012).

However, some aspects of forming an integrated knowledge network remain challenging for students across courses. This was demonstrated in our results showing no significant differences between introductory and upper-division student populations' ability to sort entities into mechanism-appropriate mental categories and connect molecular mechanisms to their roles in the cell. The fact that conflation of categories persists among the population of upper-division students suggests that the creation and modification of mental categories, a process by which knowledge about the central dogma is organized, is continuous and ongoing as education proceeds. Although our data cannot suggest patterns in knowledge integration over time or make hypotheses about potential learning progressions, we do suggest that longitudinal studies are required to track how sorting, connecting, and integration of ideas change over time for individual students. Further, while it might be attractive to think of our proposed model as steps along a learning progression, our data do not suggest this as a linear set of steps that students must go through to gain expertise.

We were able to track a small number of students from their introductory courses to upper-division courses, because a few students volunteered for interviews in both their introductory and upper-division courses. Examination of transcripts from these students suggest, as you might expect, that some students showed significant improvement in the way they connected ideas between interviews conducted in the introductory course and interviews conducted in the upper-division course. However, we also saw evidence to suggest that some students, while showing some improvements in certain areas, also developed significant misconceptions between introductory and upper-division courses that led to conflation of DNA replication, transcription, and translation. These examples hint at the nonlinear nature of learning, in which many students may take overarching, productive steps in building scientifically normative knowledge networks, but the day-to-day restructuring and reorganizing of ideas is in fact a complex and dynamic process. Future studies will explore the impact of these ideas on conceptual change over time.

Instructional Implications

Our results suggest that some students have difficulties with appropriately differentiating mental categories for the biological mechanisms of DNA replication, transcription, and translation. Traditionally, in introductory courses and some upper-division courses, these three mechanisms have been taught within a narrow window of time, often an intentional design with the instructional objective of comparing and contrasting these similar processes. However, we suggest that instructional strategies that embed these three mechanisms within "big picture" contexts for each mechanism could be more effective in helping students to develop functional modules as mental organizers. By first creating a "need to know," students may be more likely to make causal connections between mitosis and DNA replication, gene

regulation and transcription, protein synthesis and the details of the mechanism of translation.

We do not consider this persistence of conceptual difficulties into upper-division courses to be surprising, as similar results have been found in other studies that compare introductory and upper-division students (Songer and Mintzes, 1994). In fact, we suggest that it provides further evidence for the previously established idea that knowledge integration takes time (Clark and Linn, 2003). Introductory courses in our study spent between 1 and 2 wk of instructional time on DNA replication, transcription, and translation; there is a brief review of these concepts in upper-division courses before students are asked to connect these ideas to related topics. Therefore, it is not surprising that incomplete differentiation of mental categories or non-normative connections are not uncommon, even among upper-division MCB majors. Instead, we view our results as additional evidence suggesting that coverage of fewer topics in undergraduate biology courses is needed if students are to develop deeper understanding.

Our findings also suggest that instructors should be careful when designing assessments on these topics. Assessments should reach into the types of connections students are making between ideas and the ways in which their ideas are organized. By testing underlying reasoning patterns, instead of learned vocabulary, we can communicate to students the importance of knowledge organization and integration in addition to the ability to recite declarative information. To be productive, assessments must move away from tasks that allow students to rearrange causal language associated with surface features of mechanisms or to create answers that sound like expertise without revealing whether underlying knowledge remains organized by associative reasoning.

The process of knowledge integration is neither linear nor straightforward. As instructors, when we observe non-normative ideas in our classrooms, it is only natural to return to the original lesson plan with a critical eye. However, our results support the idea that student knowledge integration is an essential, internal, and individual process by which students need to actively reorganize and reinterpret acquired knowledge. As a result, when assessments reveal non-normative ideas, we must be cognizant of what the particular non-normative ideas might reveal about underlying connections and organization in the students' knowledge structures. We must be aware that these non-normative connections often are not simply due to a lack of detail or effort on the part of the student but may reflect active restructuring of knowledge that sometimes creates inappropriate connections. These misconceptions may result in ideas that are commonly referred to as "misconceptions." However, the reasoning patterns that result in these misconceptions may actually represent a misapplication of reasoning that could lead to a productive connection in another context. We urge instructors to cultivate intentional awareness of how students are sorting ideas, creating connections, and integrating knowledge, as these are integral to the dynamic and continual process of learning.

ACKNOWLEDGMENTS

We acknowledge Dr. Phillip Chapman, Professor Emeritus, Colorado State University, for his guidance in our statistical analysis.

REFERENCES

- Bolger MS, Kobiela MA, Weinberg PJ, Lehrer R (2012). Children's mechanistic reasoning. *Cogn Instr* 30, 170–206.
- Brown DE (2013). Students' conceptions as dynamically emergent structures. *Sci Educ* 23, 1463–1483.
- Chase WG, Simon HA (1973). Perception in chess. *Cognitive Psychol* 4, 55–81.
- Chi MT (2013). Three types of conceptual change: belief revision, mental model transformation, and categorical shift. In: *International Handbook of Research on Conceptual Change*, ed. S. Vosniadou, New York: Routledge, 61–82.
- Chi MT, Feltovich PJ, Glaser R (1981). Categorization and representation of physics problems by experts and novices. *Cogn Sci* 5, 121–152.
- Chi MT, Koeske RD (1983). Network representation of a child's dinosaur knowledge. *Dev Psychol* 19, 29–39.
- Clark D, Linn MC (2003). Designing for knowledge integration: the impact of instructional time. *J Learn Sci* 12, 451–493.
- Clark D, Linn MC (2013). The knowledge integration perspective: connections across research and education. In: *International Handbook of Research on Conceptual Change*, ed. S. Vosniadou, New York: Routledge, 520–538.
- Craver CF (2001). Role functions, mechanisms, and hierarchy. *Philosophy Sci* 68, 53–74.
- Craver CF (2002). Structures of scientific theories. *Blackwell Guide to the Philosophy of Science* 19, 55–79.
- Crick F (1970). Central dogma of molecular biology. *Nature* 227, 561–563.
- Darden L (2002). Strategies for discovering mechanisms: schema instantiation, modular subassembly, forward/backward chaining. *Philosophy Sci* 69, 354–365.
- Darden L (2008). Thinking again about biological mechanisms. *Philosophy Sci* 75, 958–969.
- De Groot AD (1965). *Thought and Choice in Chess*, The Hague: Mouton.
- diSessa AA (1988). Knowledge in pieces. In: *Constructivism in the Computer Age*, ed. G Forman and PB Pufall, Hillsdale, NJ: Erlbaum, 49–70.
- diSessa AA (1993). Toward an epistemology of physics. *Cogn Instr* 10, 105–225.
- Duncan RG (2007). The role of domain-specific knowledge in generative reasoning about complicated multileveled phenomena. *Cogn Instr* 25, 271–336.
- Duncan RG, Reiser BJ (2007). Reasoning across ontologically distinct levels: students' understandings of molecular genetics. *J Res Sci Teach* 44, 938–959.
- Duncan RG, Tseng KA (2011). Designing project-based instruction to foster generative and mechanistic understandings in genetics. *Sci Educ* 95, 21–56.
- Gigerenzer G, Gaissmaier W (2011). Heuristic decision making. *Annu Rev Psychol* 62, 451–482.
- Gilbert JK, Osborne RJ, Fensham PJ (1982). Children's science and its consequences for teaching. *Sci Educ* 66, 623–633.
- Gobbo C, Chi M (1986). How knowledge is structured and used by expert and novice children. *Cogn Dev* 1, 221–237.
- Guindon R (1990). Knowledge exploited by experts during software system design. *Int J Man Mach Stud* 33, 279–304.
- Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999). From molecular to modular cell biology. *Nature* 402, C47–C52.
- Hofmann KP, Spahn CM, Heinrich R, Heinemann U (2006). Building functional modules from molecular interactions. *Trends Biochem Sci* 31, 497–508.
- Jonassen D (2000). Toward a design theory of problem solving. *Educ Tech Res Dev* 48, 63–85.
- Lehrer R, Schauble L (1998). Reasoning about structure and function: children's conceptions of gears. *J Res Sci Teach* 35, 3–25.
- Linn M (1995). Designing computer learning environments for engineering and computer science: the scaffolded knowledge integration framework. *J Sci Educ Tech* 4, 103–126.
- Machamer P, Darden L, Craver CF (2000). Thinking about mechanisms. *Philosophy Sci* 67, 1–25.
- Marbach-Ad G, Stavy R (2000). Students' cellular and molecular explanations of genetic phenomena. *J Biol Educ* 34, 200–205.
- McClary L, Talanquer V (2011). Heuristic reasoning in chemistry: making decisions about acid strength. *Int J Sci Educ* 33, 1433–1454.
- Metz KE (1985). The development of children's problem solving in a gears task: a problem space perspective. *Cogn Sci* 9, 431–471.
- National Research Council (2000). *How People Learn*, Washington, DC: National Academies Press.
- Özdemir G, Clark DB (2007). An overview of conceptual change theories. *Eurasia J Mathematics Sci Technol Educ* 3, 351–361.
- Russ RS, Scherr RE, Hammer D, Mikeska J (2008). Recognizing mechanistic reasoning in student scientific inquiry: a framework for discourse analysis developed from philosophy of science. *Sci Educ* 92, 499–525.
- Songer CJ, Mintzes JJ (1994). Understanding cellular respiration: an analysis of conceptual change in college biology. *J Res Sci Teach* 31, 621–37.
- Southerland SA, Abrams E, Cummins CL, Anzelmo J (2001). Understanding students' explanations of biological phenomena: conceptual frameworks or p-prims? *Sci Educ* 85, 328–348.
- Van Mil MH, Boerwinkel DJ, Waarlo AJ (2013). Modelling molecular mechanisms: a framework of scientific reasoning to construct molecular-level explanations for cellular behaviour. *Sci Educ* 22, 93–118.
- Venville GJ, Treagust DF (1998). Exploring conceptual change in genetics using a multidimensional interpretive framework. *J Res Sci Teach* 35, 1031–1055.
- Vosniadou S (2013a). Conceptual change in learning and instruction: the framework theory approach. In: *International Handbook of Research on Conceptual Change*, ed. S. Vosniadou, New York: Routledge, 11–30.
- Vosniadou S (ed.) (2013b). *Educational Psychology Handbook: International Handbook of Research on Conceptual Change*, 2nd ed., Florence, KY: Routledge.
- Vosniadou S, Brewer WF (1992). Mental models of the earth: a study of conceptual change in childhood. *Cogn Psychol* 24, 535–585.