Received: 30 April 2017 | Revised: 16 April 2018 | Accepted: 20 April 2018
DOI: 10.1002/tea.21466

RESEARCH ARTICLE

WILEY | JRST

Categories in conflict: Combating the application of an intuitive conception of inheritance with category construction

Joshua Premo[1] | **Andy Cavagnetto**[1,2] | **Garrett Honke**[3] | **Kenneth J. Kurtz**[3]

[1] School of Biological Sciences, Washington State University, 297 Eastlick Hall, Pullman, Washington

[2] Department of Teaching and Learning, Washington State University, 335 Cleveland Hall, Pullman, Washington

[3] Department of Psychology, Binghamton University, PO Box 6000, Binghamton, New York

Correspondence
Joshua Premo, 1125 NE Washington St, Pullman, NY.
Email: joshua.premo@wsu.edu

Funding information
IES Cognition and Student Learning, Grant/Award Number: R305A120554

Abstract

The idea that characteristics acquired by an organism during its lifetime can be inherited by offspring and result in evolution is a substantial impediment to student understanding of evolution. In the current study, we performed a preliminary examination of how acquiring physical changes in a question prompt may differentially cue intuitive and scientific justifications of inheritance and evolution and how this varies based on how student learned the concept. Middle school students in a suburban northeastern district ($N = 314$) either learned about evolutionary change with a category construction task (with different levels of feedback support) or completed a worksheet. Three days later students responded to two free response scenarios (one where a physical change is acquired). Responses were coded based on student justifications for either science accuracy or intuitive nature. Specific reasons were coded by justification type with high inter-rater agreement ($k > 0.93$). Results showed that students were more likely to apply intuitive reasoning when a physical change was acquired (50%) than if the change was behavioral in nature (16%). Additionally, students who completed the category construction task provided significantly more scientifically accurate justifications about inheritance ($M = 1.12$) than control students ($M = 0.47$), and significantly less intuitive justifications ($M = 0.67$) than control ($M = 1.13$). Finally, category construction produced the most scientific reasoning when feedback was provided. Taken together, these results suggest that intuitive reasoning is differentially applied based on physical organismal changes, intuitive reasoning is less frequent when learning via category construction, and the category construction task is more effective for this population with the inclusion of feedback.

KEYWORDS
categorization, evolution learning, intuitive conception, middle school, misconception

1 | INTRODUCTION

Understanding the process by which educators can encourage students to apply scientifically accepted explanations of phenomenon over incorrect intuitive ideas[1] is a central challenge of science education. The application of incorrect ideas related to scientific phenomena—including intuitive conceptions—is widespread and often highly resistant to change. Significant efforts have been made to track how ideas change during the process of learning science. Many different interpretations of how student ideas change have been noted with claims being made that conceptions become systematically organized (diSessa, 1993), restructured (Vosniadou, 1992), abandoned (de Posada & María, 1997), corrected (Chen, Pan, Sung, & Chang, 2013), or go through other forms of alteration (Heddy & Sinatra, 2013) in the process of learning. Given that students tend to apply fewer intuitive ideas during the process of learning science (Abraham, Perez, Downey, Herron, & Meir, 2012), it is unsurprising that improving understanding via altering intuitive ideas is a widely accepted goal when attempting to decrease the prevalence of scientific misconceptions.

An alternative hypothesis, however, is that at least some intuitive ideas are not changed during learning, but instead students learn to inhibit the application of intuitive ideas in favor of scientifically accurate understandings. In other words, the process of learning science may be, in part, a process of establishing contextual bounds by which intuitive ideas and their scientifically correct counterparts can be applied. This state—in which students simultaneously have both fully formed intuitive and scientific understandings of a given conception and apply them selectively based on context—has been defined as coexistence (Gelman, 2011; Legare & Visala, 2011; Mortimer, 1995; Shtulman & Lombrozo, 2016).

Yet, not all intuitive ideas may have equal ability to retain this degree of permanence. Coexistence should have a higher likelihood when abandoning an intuitive idea leads to a disadvantage outside of the classroom. For example, being able to distinguish between the genotype (genetic information) and phenotype (physical characteristics) of an organism is fundamental to understanding evolution and genetics, yet students are particularly resistant to differentiating these concepts (Lewis & Kattmann, 2004; Marbach-Ad & Stavy, 2000; Venville, Gribble, & Donovan, 2005). This may result in students conflating the physical features of an organism with its genetic composition, and predispose them to *incorrectly* accept that physical changes accumulated during life (e.g., strength) will be passed on to offspring. The developmental finding that children intuitively understand inheritance through physical features (Kaminski, Gentaz, & Mazens, 2012; Williams, 2013; Williams & Smith, 2010) suggests that the source of this misconception is intuition, and thus unlikely to be erased via science instruction.

If conceptual change is unlikely to occur in this case, then focusing on how instructors can encourage students to selectively apply science knowledge of inheritance while reducing the use of intuitive ideas is of primary concern. To explore this idea, the present study sought to answer the following questions: (a) To what extent do seventh grade students apply intuitive ideas of inheritance when a physical change is present (Scenario 2) versus when a behavioral change is present (Scenario 1)? (b) Do students who learn about inheritance in an evolutionary context via a category construction task demonstrate more scientific reasoning and less intuitive reasoning than a worksheet control? Based on past research on coexistence and the examination of student misconceptions discussed above, we expect to find coexistence of science-based and intuitive reasoning as evidenced by a differential use of these sources of reasoning between scenarios. If category construction is an effective tool for promoting the application of science-based knowledge, explanations of evolutionary change should rely less on intuition and more on scientific knowledge after the category construction learning task.

2 | CONCEPTUAL COEXISTENCE

Student's at all academic levels hold intuitive conceptions that are incompatible with scientific reasoning. Coexistence between intuitive and scientifically accurate ideas has been repeatedly noted in science misconception research. At the primary level, Schneider and Hardy (2013) looked at student understanding of sinking and floating among third grade students at three different time points. They found that student understanding included intuitive conceptions about floating and sinking that co-existed with science-based reasoning and persisted throughout the study. At the secondary and tertiary level, Thorn, Bissinger, Thorn, and Bogner (2016) examined knowledge of tree assimilation and wood synthesis among sixth and tenth grade students, and freshman college students (both science and non-science majors). The resulting cluster analysis of student response patterns showed that, despite differences in exposure to scientific knowledge between these groups, both intuitive conceptions and scientific understanding were simultaneously present in most student response patterns. A similar result was found in Clark's (2006) longitudinal (Grades 8 to 12) interviews. Students exhibited both intuitive and scientific ideas about thermal equilibrium in six out of seven interviews. Conceptual coexistence was also found for heat and temperature knowledge at all seven time-points. This coexistence of contradictory knowledge has also been found in experimental work examining the ability of children and adults to learn selection-based explanations of biological adaptation (Shtulman, Neal, & Lindquist, 2016). Results showed that while older children and adults could equally learn these explanations, both groups included intuitive ideas alongside of scientifically accepted ideas. These studies provide evidence that contradictory explanatory frameworks are simultaneously held by students over extended periods of time, suggesting prevalent and long-term effects of conceptual co-existence on scientific reasoning. These findings are consistent with psychological research on simultaneous contradictory belief (Sloman, 1996) and dual-system accounts of human reasoning that posit one system that is fast and intuitive and another that is slow and strategic (e.g., Evans, 2008; Evans & Stanovich, 2013; Kahneman, 2011).

The use of intuitive and scientifically sound reasoning and conceptual knowledge appears to be affected by the context of the assessment. For example, our own work has found that scientific situations can decrease the use of particular logical fallacies when compared to everyday contexts (Cavagnetto & Kurtz, 2016). Similarly, students differentially access intuitive conceptions of evolution depending on the type of organism present in assessment items (Ha, Lee, & Cha, 2006). Students have also been shown to provide conflicting answers on thermal equilibrium questions based on contextual differences in interviews (Clark, 2006). It has even been found that students will apply different ideas to the same physics concept based on presentation context (Bryce & MacMillan, 2009; Sabella & Redish, 2007). Surprisingly nuanced contextual features of evolution knowledge (e.g., trait loss versus gain) can also affect the retrieval and use of intuitive conceptions by students (Nehm, Beggrow, Opfer, & Ha, 2012; Nehm & Ha, 2011). These results demonstrate that students can and often do maintain alternative (and contradictory) conceptual knowledge and apply it selectively based on the context. Furthermore, the finding that students can apply a scientifically correct idea in a specific classroom context, but revert to incorrect ideas in a new or less familiar situation (Oliver, 2011) suggests that knowledge can be inert outside of the classroom context (see Barnett & Ceci for review, 2002). While learning can predispose students to apply science-based knowledge in classroom instruction contexts, transfer to even slightly dissimilar contexts can be incredibly difficult (Pugh, Koskey, & Linnenbrink-Garcia, 2014). Spontaneous transfer of prior knowledge in novel domains is notoriously difficult and the problem is compounded if it requires a student to override an intuitive conception that has value in everyday thinking.

3 | INTUITIVE CONCEPTIONS OF EVOLUTION

Evolution education is a fruitful area for research on promoting science-based reasoning. Previous attempts at targeted instructional intervention show minimal change (Bishop & Anderson, 1990; Demastes, Settlage, & Good, 1995; Gregory, 2009; Smith, 2010). Understanding of evolutionary theory is alarmingly low in the United States (Alters & Nelson, 2002; Miller, Scott, & Okamoto, 2006); even science graduate students can have persistent misunderstandings of the subject (Gregory & Ellis, 2009). Given the attention that learning of evolution and related misconceptions have received, student difficulties are unlikely to be a result of either research neglect or a lack of instructional effort. Rather it may be that intuitive conceptual knowledge perseveres in the face of scientific instruction due to existing cognitive biases (Coley & Tanner, 2012; Shtulman & Calabi, 2012; Sinatra, Brem, & Evans, 2008) including teleology (Kampourakis, Palaiokrassa, Papadopoulou, Pavlidi, & Argyropoulou, 2012; Kelemen, 2012; Tamir & Zohar, 1991), essentialism (Emmons & Kelemen, 2015; Gelman & Rhodes, 2012; Shtulman & Schulz, 2008), and anthropomorphism (Legare, Lane, & Evans, 2013; Tamir & Zohar, 1991).

These cognitive biases may predispose individuals toward scientific explanations that assume that: (1) evolution has a goal and purpose (teleological reasoning) (Kelemen, 1999), (2) entities with like properties are a single category with deep commonalities (essentialist reasoning) (Haslam & Whelan, 2008), and (3) human characteristics can be used to make inferences about nonhuman species or entities (anthropomorphic reasoning) (Coley & Tanner, 2012). If these intuitive conceptions are addressed in the context of everyday nonscientific reasoning, however, they hold value. Teleological reasoning can help to reason about why events occur (Dennett, 1989). Essentialist thinking can promote inductive generalization or abstraction of new concepts (Gelman, 2003). Anthropomorphizing human preferences and desires for food, shelter and kin health can be useful for understanding the behavior of nonhuman animals. These examples demonstrate that intuitive conceptual knowledge has utility for human cognitive functioning. It cannot simply be viewed as the less-scientific precursor to more accurate understanding of scientific phenomena.

4 | INTUITIVE UNDERSTANDING OF INHERITANCE

Students develop intuitive understandings of inheritance during childhood long before they are exposed to ideas of genetic inheritance in the classroom (Kaminski et al., 2012). Conflict between these intuitive conceptions and scientific views of inheritance can be seen in students perceiving genes as de facto physical traits (Venville et al., 2005), as “small, trait-bearing, particles” (Lewis & Kattmann, 2004, p. 195), or as an “active particle that controls characteristics” (Venville & Treagust, 1998, p. 1031). Student knowledge of inheritance exhibits a common bias in understanding that genetic materials are analogous to physical characteristics (e.g., Marbach-Ad, 2001; Marbach-Ad & Stavy, 2000), a misconception that is representative of early scientific models of inheritance (Gericke & Hagberg, 2007). Students perceive genotype and phenotype as being at the same hierarchical level and this promotes a focus on using observable characteristics of organisms to understand their genotype (Lewis & Kattmann, 2004).

TABLE 1 Origins and persistence of the merged conceptualization of physical traits and genetic make-up

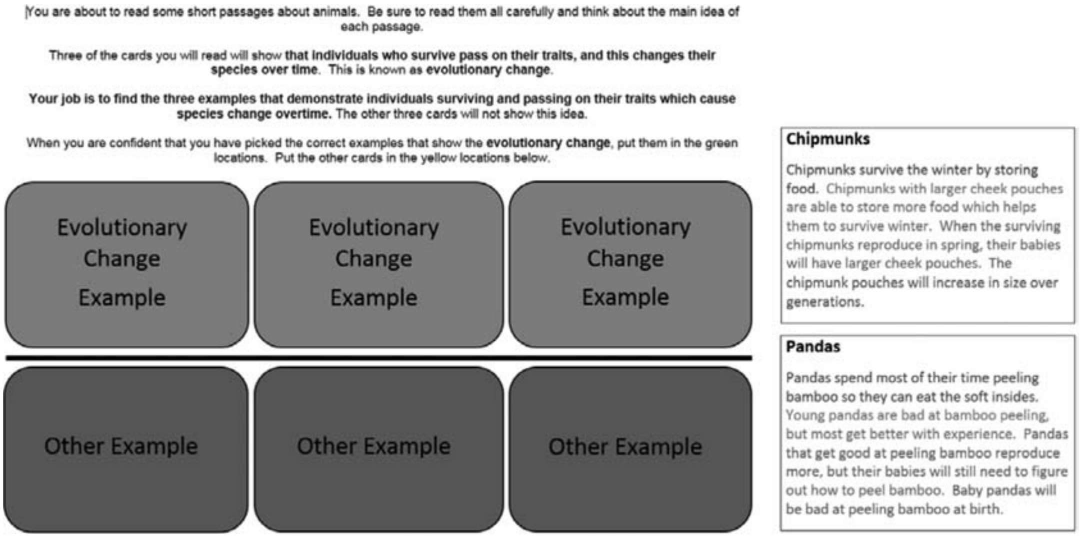
Intuitive Conception	Origins in development	Student views in the classroom
Physical traits of an organism are the same as its genetic composition.	• Early in life children use physical features of parents to determine inheritance (Williams, 2012; Williams & Smith, 2010). • Children by the age of five automatically use physical facial resemblance to accurately judge genetic kinship (Kaminski et al., 2012).	• Genes are de facto physical traits (Venville et al., 2005). • Genes are "small, trait-bearing, particles" (Lewis & Kattmann, 2004, p. 195). • Genes are "active particles that control characteristics" (Venville & Treagust, 1998, p. 1031). • Genetic materials are analogous to physical characteristics (Marbach-Ad, 2001; Marbach-Ad & Stavy, 2000).

Children intuitively understand inheritance through physical features (Kaminski et al., 2012). By age four, children intuitively understand that physical features of parents are inherited by their offspring (Williams, 2012; Williams & Smith, 2010). By age five, children will automatically use physical facial resemblance to accurately judge genetic kinship (Kaminski et al., 2012) suggesting an intuitive origin for using physical properties to infer biological kinship. We speculate that intuitive conceptual knowledge promotes the idea that physical (phenotypic) change can be inherited by offspring (Table 1). This contributes to the observed resilience of misconceptions about evolutionary change. In this study, we explore this hypothesis by examining differences in science-based and intuitive reasoning patterns about inheritance in two scenarios. One prompt exhibited a behavioral change and the other included a behavior which resulted in a physical change. It was hypothesized that intuitive knowledge of inheritance processes would predispose students to incorrectly claim that inheritance will take place in the physical change scenario more often than in the scenario with the behavioral change alone.

5 | CATEGORY CONSTRUCTION

The comparison of examples that share an underlying principle is an effective tool for promoting learning and transfer across many domains of study (see for review, Alfieri, Nokes-Malach, & Schunn, 2013). Comparison learning is particularly well-studied and effective for improving science learning outcomes (Chen & Klahr, 1999; Glynn, 1991; Klahr & Li, 2005; Kurtz, Boukrina, & Gentner, 2013; Kurtz, Miao, & Gentner, 2001; Matlen & Klahr, 2013). Considering the extensive laboratory evidence that humans have difficulty retrieving useful knowledge in new or distinct contexts (Gentner, Loewenstein, Thompson, & Forbus, 2009; Gentner, Rattermann, & Forbus, 1993; Gick & Holyoak, 1980, 1983; Loewenstein, Thompson, & Gentner, 1999), an instructional tool that can improve this performance could have profound impact on education outcomes—particularly within the domain of science instruction.

In the category construction task, learners are asked to sort a set of unlabeled examples into coherent groups that instantiate a target scientific principle and counter examples. This technique builds on comparison-based learning by integrating an additional component of unsupervised category learning. In the present experiment, a set of positive examples (and foils) were sorted that target the principle of evolutionary change—organisms passing on their genes and increasing the presence of a trait in a population. The goal of the task is to identify which passages exemplify the target principle (evolutionary change) and then group them based on their adherence to that principle (see Figure 1). Through this activity, the principle is learned as a *relational* category—a category defined by relations among entities



You are about to read some short passages about animals. Be sure to read them all carefully and think about the main idea of each passage.

Three of the cards you will read will show **that individuals who survive pass on their traits, and this changes their species over time**. This is known as **evolutionary change**.

Your job is to find the three examples that demonstrate individuals surviving and passing on their traits which cause species change overtime. The other three cards will not show this idea.

When you are confident that you have picked the correct examples that show the **evolutionary change**, put them in the green locations. Put the other cards in the yellow locations below.

Chipmunks

Chipmunks survive the winter by storing food. Chipmunks with larger cheek pouches are able to store more food which helps them to survive winter. When the surviving chipmunks reproduce in spring, their babies will have larger cheek pouches. The chipmunk pouches will increase in size over generations.

Pandas

Pandas spend most of their time peeling bamboo so they can eat the soft insides. Young pandas are bad at bamboo peeling, but most get better with experience. Pandas that get good at peeling bamboo reproduce more, but their babies will still need to figure out how to peel bamboo. Baby pandas will be bad at peeling bamboo at birth.

FIGURE 1 Sort mat from the Category Construction task (left) and sample scenarios (right). The top scenario (chipmunks) is an example of the principle consistent (evolutionary change) scenarios. The bottom scenario (pandas) is a counterexample. Note: The counterexample includes explicit offspring noninheritance of the behavioral change

(e.g., carnivore)—that can be used to identify instances of a scientific principle and make sense of the world (Gentner & Kurtz, 2005). Category-based learning is critical for education; relational categories are particularly ubiquitous in science education (e.g., force, catalyst, predator) (Goldwater & Schalk, 2016). Relational categories form a key basis of knowledge organization for experts (Chi, Feltovich, & Glaser, 1981). In this case, we hypothesize that learners will be better equipped to identify evolutionary change and retrieve science-based knowledge when reasoning about ecological situations and even situations that are relationally similar but lack an ecological or biological context (e.g., artificial evolutionary processes such as evolutionary algorithms, for an example see Lamb & Premo, 2015).

The process of categorizing examples and nonexamples of a principle as a tool for learning has its theoretical roots in structure mapping theory (Gentner, 1983)—the idea that the alignment and correspondence of structured, conceptual representations is a key requisite for analogical reasoning and similarity judgment. Research inspired by this theoretical framework has shown that human memory retrieval will most frequently produce nonrelational remindings, even though relationally similar knowledge has more value for higher-order cognitive processes like comparison, similarity judgment, and analogical reasoning (Gentner et al., 1993; Gick & Holyoak, 1980). The key hypothesis underlying our use of the category construction task is that the process of treating superficially distinct but relationally similar examples as members of a category can mitigate the well-studied failure of relational memory retrieval (Kurtz & Honke, under review). In other words, we aim to promote the retrieval and transfer of science-based knowledge by teaching classroom principles as categories (see also Goldwater & Schalk, 2016). We hypothesize that principles that have *category status* will be accessible and retrieved as fluidly as attribute-based categories: natural kinds (e.g., dogs, trees) and artifacts (e.g., chairs, cars). Simply stated, scientific principles learned by exposure to multiple examples in the category construction task should be easier to identify when encountered in a manner akin to how we fluently access attribute-based categories like "truck."

Category construction builds on the importance of schema formation by seeking to optimize transfer by promoting the establishment of category status. Schema formation promotes transfer by establishing a link between a specific knowledge cue and the abstracted schema within long term memory (Gentner et al., 1993; Loewenstein, 2010). Category construction seeks to promote the establishment of more effective recall and transfer by investigating ways in which abstract schemas can be formalized into a generalized representation (encompassing at least several cases) made more discrete via contrast with nonexamples of a given category. Laboratory studies investigating the category status have recently began, with the first evidence that category construction is effective for promoting learning and transfer having recently been evidenced (Kurtz & Honke, under review).

The present work is designed to investigate how category construction can be integrated into authentic classroom settings and to what degree it can promote the use of science-based conceptual knowledge over alternatives; in this case intuitive student understanding of inheritance. There are two main research questions in this work: To what extent do seventh grade students apply intuitive understandings of inheritance when a physical change is present (Scenario 2) versus when a behavioral change is present (Scenario 1)? and Do students who learn about inheritance in an evolutionary context via category construction demonstrate more scientific reasoning and less intuitive reasoning than a worksheet control? Additionally, a subgoal of this work was to investigate how varying levels of instructor feedback support learning in the category construction task. Based on the research previously discussed, we expect to find the coexistence of science-based reasoning and intuitive reasoning; we expect this reasoning to be differentially used based on the presence of physical versus behavioral changes in the assessment prompts. If the category construction task is indeed effective for promoting the application of science-based knowledge, explanations of evolutionary change should rely less on intuitive conceptual knowledge and more on science knowledge in this condition. After answering these questions, we performed a secondary analysis to examine the extent to which receiving the correct answer to a category construction task (a basic form of feedback) contributed to the effectiveness of the task.

6 | METHOD

6.1 | Participants

As part of a larger study looking at the impact of category construction on student understanding and transfer of evolutionary ideas, seventh grade science students ($N = 314$) from a middle-sized district in the Northeastern United States participated in the experiment as part of their classroom instruction. The demographics of the school include 48% female, 79% Caucasian, 7% African American, 8% Hispanic, 2% Asian or Pacific Islander, and 7% multiracial. Students with disabilities compromised 14% of the school's population, and 51% of the students were eligible for free or reduced lunch.

6.2 | Context

Three teachers were responsible for teaching all of the life science classes in the middle school where the present study took place. Each teacher had prior experience teaching the class (at least three years) and taught based on the state's science standards. While each teacher demonstrated slightly different approaches to instruction in the classroom, they coordinated with one another and used many of the same activities and focused on all of the same concepts at approximately the same time (within a day or two of instruction). The current intervention took place in the middle of the evolution unit. Students had finished learning the multiple forms of evidence for evolution and were beginning a laboratory activity on natural selection. Students completed either the category construction task or an information-matched worksheet at the end of class on a Friday and were assessed at the start of class on the following Monday. Students had no formal exposure to genetics or explicit instruction on inheritance in their classroom prior to the study.

6.3 | Design and procedure

Classes were pseudo-randomly assigned to the category construction or worksheet control conditions based on the constraint that each condition received classrooms with similarly rated academic ability (as assessed by the teachers). There were three category construction conditions designed to assess the effect of feedback: no feedback ($n = 44$), individualized feedback ($n = 71$), and classroom feedback ($n = 53$); and the worksheet control condition ($n = 53$). Feedback is defined as students receiving information about the correct answer to the categorization task with explanation following task completion. Students who did not receive feedback simply completed the task with no follow-up or other supports. Initial analysis found that there were no significant differences among these category construction conditions in terms of either the average number of scientific conceptions or intuitive conceptions students used. Thus, the category construction conditions were collapsed for the purposes of research questions one and two. Even though category construction conditions did not demonstrate significant differences in reasoning relative to one another, they could be differentially related to worksheet student reasoning. Therefore, a secondary analysis was performed (RQ#3) to examine the extent to which each category construction condition differed from the worksheet control.

Students in the category construction task received a sort mat (Figure 1) presenting the target evolutionary principle: *"individuals who survive pass on their genetic traits, and this changes their species over time."* The instructions were to *"find three examples that demonstrate individuals surviving and passing on their genetic traits – leading to species change over time."* Students were provided with six scenarios printed on laminated cards. These scenarios featured sentences that were color-coded such that the students could identify the corresponding sentences in each scenario. The color coding scheme was the result of a series of norming studies that indicated—in contrast to college-level investigations—that middle school science students need additional support identifying correspondences between scenarios (Honke et al., 2015; Kurtz et al., 2014). Three of the provided scenarios represented the target principle and three counterexamples presented organisms learning a behavior during their lifetime that did not result in evolutionary change (referred to as "other example" on the sort mat). The counterexamples were designed to highlight the idea that behavioral change is not passed to offspring through heredity as an *alignable difference* from the principle consistent examples (Markman & Gentner, 1996, 1997). This contrast was used to highlight and address a documented misconception that students have about heredity (Lewis & Kattmann, 2004; Marbach-Ad, 2001; Marbach-Ad & Stavy, 2000).

The worksheet control group received a one-page worksheet that presented the target evolutionary principle and the three principle-consistent scenarios provided in the category construction task (the non-evolutionary change counterexamples were not provided). The worksheet featured four self-explanation questions that probed knowledge of the principle of heredity and evolutionary change (see supplementary materials). It was informationally consistent in terms of the target principle and scenarios. The worksheet employed self-explanation, an alternative learning technique favored for fostering relational knowledge like science principles (Williams & Lombrozo, 2010, 2013). Lastly, its structure was commonly used in these specific classroom settings. An examination of worksheets typically provided by the teachers revealed that contrast cases were not a component. Thus, contrast cases were not included in the worksheet condition to maintain alignment with the authentic instructional practices of these classrooms.

The following class day (three-day delay), all students completed an assessment composed of scenario-based free response questions (see below). The two free response prompts were designed to assess student understanding that intragenerational changes made by an individual are not inherited by their offspring. A key difference between scenarios is whether the intragenerational behavioral change had a physical/phenotypic impact. Given research that shows students typically conflate genes with physical features (Lewis & Kattmann, 2004) and intuitively judge relatedness by physical features

TABLE 2 Comparison of scenario prompts

	Scenario 1	Scenario 2
Similarities (bold) • New behavioral change • Survival advantage • Asking about the impact on later generation(s)	Some of the deer **have *developed* the habit of drinking alongside the river** rather than standing in the river, **so their feet do not get *wet and cold***. Will this **have any effect on the *offspring*** of the deer? Explain your answer.	A group of bears **have *begun* using rocks to sharpen their claws**. This makes it **easier for them to get *food***. Do you expect anything to change about the claws of these bears **across many *generations***? Explain your answer.
Differences[a] (*italics*) • Wording of the change • Survival value of change • Time frame of change		

[a] We recognize that, while necessary to make the prompts distinct, some differences between scenarios could also elicit inheritance idea in students. These are addressed at the end of the results section.

(Kaminski et al., 2012), the inclusion of a physical change was a compelling lure—likely to evoke incorrect explanations based on nonscientific conceptual knowledge (see scenarios in Table 2).

The third author collected student responses and blinded them prior to providing them to the primary author. Code categories were developed by the primary author (currently in a biology department and has completed graduate coursework in microgenetics and evolutionary theory) to classify student justifications *within* responses as scientific, intuitive, or other. This scheme was informed by a diverse body of research on evolution and inheritance misconceptions (See above section on Intuitive Concepts of Inheritance, p. 8) and sought to capture the majority of justifications provided by students. The primary author and secondary author (science education faculty member with appointment in both biology and education departments) then coded all justifications within student responses independently from one another. Results were compared after independent coding of all student justifications. This revealed acceptable agreement during the initial classification ($k > 0.93$ per category, see Table 1) and differences were resolved by discussion to reach 100% agreement and generate the final dataset.

Two aspects of the responses were examined in the coding scheme: (1) whether the student indicated that the behavioral change would have an impact on subsequent generations and (2) the justification(s) provided for this position. While (1) is a dichotomous yes/no rating, students could provide multiple justifications for their position in response to the prompt. Each justification was included in the analysis of both students reasoning between scenarios, as well as between category construction and control groups. Once the justifications were identified, each was judged to be science aligned, intuitive, or other. Science aligned justifications were defined by congruence with evolutionary theory. This included arguing for no impact, justifications based on learned behavior, recognition of the greater likelihood of survival by the parent generation (thus, making them better able to raise young), and answers referencing genetics or no heritability. Intuitive justifications included reasoning for significant impacts on future generations that either explicitly included inheritance (e.g., "Yes because when two deer *reproduce the younger deer will get the traits*") or implied inheritance (e.g., "Yes they will become sharper *with more generations*"). Justifications based on additional physical traits in response to the scenario were also included in this category. For more examples of codes that were considered science aligned and intuitive and for student examples of each category see Table 3.

TABLE 3 Student reasoning codes

Impact	Justification	Type	Example for S1	Example for S2
No	They wouldn't know the behavior	Sci	"No because this is not a trait the baby *will have to learn* this."	"The offspring's parents might teach them how to, but the newborn *will not know at first*."
	Change isn't important enough to matter	n/a	"Some deer stand in the river and it doesn't change the offspring so *it makes no difference*."	"No I think even if bears has sharper claws *they wouldn't be able to sneak up on prey*."
	Change is not heritable	Sci	"No, because a deer makes the decision to go in the water it's *not a trait that can be passed on*."	"No because bears were learning something… *which is not passed on*."
	Change is just a choice/not fixed	Sci	"No this will not have an effect… *because the deer choose* to stand on the side of the river."	"No because *they are doing this by themselves*, it's not a trait."
	Survival reasoning (doesn't impact survival)	Sci	"No because if their feet are cold *they still have the ability to find a mate and reproduce*."	"No because if their claws *already get them food* they don't need to change."
	The animal was not born with the change	Sci	"The deer *were not born* with this skill or intelligence."	"No because… *they were not born* with the claws that they are getting from sharpening them."
	Change is not genetic/DNA based	Sci	"This will not have any effect on the offspring of the deer because it is just a habit *not a genetic trait*."	"No the bear's offspring's claws would not change because sharpen claws are *not in their DNA*."
Yes	Survival reasoning (change increases survival)	Sci	"Yes *more deer will survive* because their feet will not be wet and cold."	"Yes the bears with sharp claws *will live and reproduce*."
	Significant future impact (implicit inheritance)	Int	"Yes because *the young deer will gain this trait* and they will begin doing this."	"Yes they *will become sharper with more generations*."
	Through learning the change	Sci	"Yes because *the offspring will learn this habit* from their parents."	"I do because the bears that aren't doing *this may imitate the bears that are*."
	Explicit inheritance of the change	Int	"Yes because when two deer *reproduce the younger deer will get the traits*."	"I think eventually the species of bears will have naturally sharp claws because *it is a trait that will be passed down to each bear's offspring*."
	Through additional physical traits	Int	"Yes I think their necks might *adapt to be long* to get water easier."	"Yes more sharp and *probably longer claws*."

Note: Letters representing categories of reasoning include "Sci" (Science aligned justification is defined by whether or not the justification is congruent with evolution), "Int" (Justifications aligned with intuitive understandings of inheritance that conflict with science) and "n/a" (Justification that the change was "not important enough" could be interpreted as other).

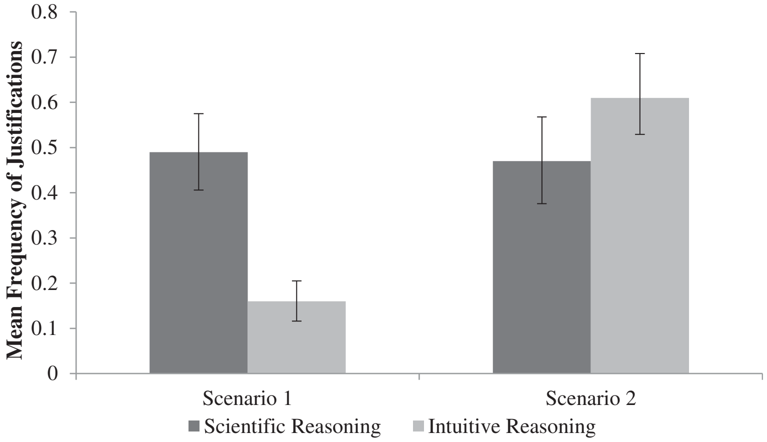


FIGURE 2 Difference in science aligned and intuitive justifications provided in each scenario. Note that that while students provided approximately the same amount of science aligned reasons regardless of scenario, there was a strong contrast between the intuitive reasons provided by students in Scenario 1 (no physical change) and Scenario 2 (physical change). Error bars represent the 95% confidence interval for the bootstrapped mean

7 | RESULTS

Prior to analysis, students who did not complete the questions or provided an illegible/nonsensical answer were removed. The remaining students provided responses to both scenarios (category construction, $n = 172$; worksheet control, $n = 53$). Performance scores on a unit test from an unrelated unit (cell biology) that immediately preceded the evolution unit ($M_{CC} = 70.4$; $M_{WC} = 70.1$) and the evolution unit test ($M_{CC} = 78.7$, $M_{WC} = 77.8$) were used to confirm that random assignment to condition produced relatively balanced groups. The groups were not significantly different in their performance on the unit tests ($ps > .5$).

Data normality was assessed using Shapiro–Wilk tests. Results of the Shapiro–Wilk tests showed significant non-normality in all variables, thus nonparametric tests were used (Field, 2013). We attempted to further account for both non-normality and unequal samples sizes with a sample-based bootstrapping approach. Primary nonparametric analyses between scenarios and conditions were bootstrapped (1,000 iterations, resampling with replacement) and cumulative results were examined to make sure that significance was still found in greater than 95% of bootstrapped samples. Results will first be presented on the differences in student reasoning across scenarios using Wilcoxon signed-rank tests (1 sample/paired) to compare the quantity and average number of reasoning types between scenarios. This is followed by McNemar's test for paired nominal data to examine more nuanced differences in each reasoning category by scenario. Finally, differences between the category construction and worksheet control groups are reported. Wilcoxon 2-Samples tests were used to compare mean frequency of intuitive and science justifications between the category construction and worksheet control groups.

RQ#1 *To what extent did seventh grade students apply intuitive understandings of inheritance when a physical change was present (Scenario 2) versus when a behavioral change was present (Scenario 1)?*

First, we examined student application of both scientific and intuitive reasoning between scenarios (Figure 2). We hypothesized that students would be more likely to apply intuitive reasoning in Scenario 2 (where a physical change took place) than Scenario 1, where only a behavioral change took place. Descriptive statistics show that 41% and 33% of students provided science-aligned responses for

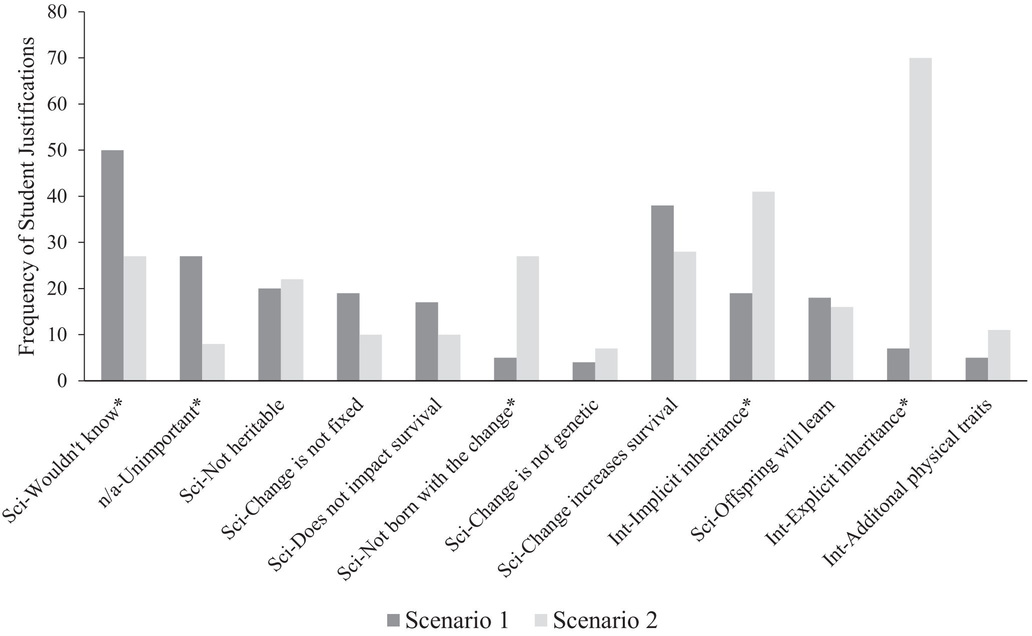


FIGURE 3 Comparison for student use of justifications between scenarios ($n = 225$). Note that students could include more than one justification per response. "Sci" (Science aligned justification is defined by whether or not the justification is congruent with evolution), "Int" (Justifications aligned with intuitive understandings of inheritance that conflict with science) and "n/a" (Justification that the change was "not important enough" could be interpreted as other). * Significant difference between scenarios (McNemar's test for paired nominal data, $p < .01$). See Table 3 for definitions and descriptions of each justification category

Scenario 1 and Scenario 2, respectively. The bootstrapped mean number of science justifications provided by students was 0.49 (95% CI 0.41–0.57) in Scenario 1 and 0.47 (95% CI 0.38–0.57) in Scenario 2. A bootstrapped Wilcoxon signed-rank analysis uncovered no reliable difference between scenarios ($p > .1$), suggesting that students did not provide more science-based responses to either scenario.

However, a different pattern emerged when examining the mean proportion of students using intuitive justifications across scenarios. Intuitive justifications were provided by 16% of students in Scenario 1% and 50% of students in Scenario 2. The mean number of intuitive justifications provided by students was 0.16 (95% CI 0.11–0.21) in Scenario 1 and 0.61 (95% CI 0.52–0.71) in Scenario 2 as seen in Figure 2. A bootstrapped Wilcoxon signed-rank test showed that every iteration produced a reliable difference between conditions, $p < .001$. Thus, students demonstrated significantly more use of intuitive reasoning when a physical change was present than when only a behavioral change was present. This aligned with our hypothesis that physical changes in an organism would produce more intuitive reasoning by students. Additionally, the differential application of intuitive reasoning between scenarios (without a change in scientific reasoning) may suggest independent application. In other words, there did not appear to be an inverse relationship between student application of scientific and intuitive reasoning.

Next, we investigated shifts in student response justification. McNemar's test for paired nominal data was used to examine differences in the frequency of student scientific and intuitive justifications between scenarios (see Figure 3 for results and Table 3 for definitions and student examples). Based on our hypothesis, we would expect that students would preferentially apply intuitive explanations in the second scenario over the first. This was supported by results showing that students provided significantly more reasoning for implicit inheritance (χ^2 (1, $N = 60$) = 9.8, $p < .001$, $\varphi = 0.21$) and explicit

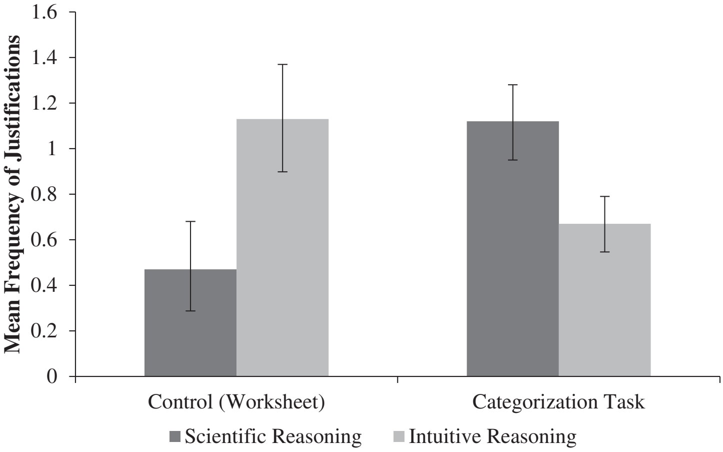


FIGURE 4 Comparison of the mean number of scientific and intuitive justifications for each condition for both scenarios combined. Error bars represent the 95% confidence interval for the bootstrapped mean

inheritance (χ^2 (1, $N = 77$) = 49.0, $p < .001$, $\varphi = 0.47$) in Scenario 2 (physical change) as compared to Scenario 1. Additional differences in student justification between scenarios were that students preferentially argued that the offspring would not know the behavior (χ^2 (1, $N = 77$) = 13.6, $p < .001$, $\varphi = 0.25$) and that the behavior was not important (χ^2 (1, $N = 35$) = 9.32, $p < .001$, $\varphi = 0.21$) in Scenario 1. Surprisingly, students preferentially reasoned that the offspring would not be born with the phenotypic change in Scenario 2 (χ^2 (1, $N = 32$) = 18.2, $p < .001$, $\varphi = 0.28$). The preferential use of noninheritance reasoning was interesting because it is a scientifically accepted idea and was preferred as a response in Scenario 2—where intuitive ideas were also more likely to be found. This phenomenon may be representative of a divergence in the ideas students are drawing on in the second scenario, with a minority recognizing that heritability was not present even with a physical change cue.

Alternative explanations for differences in student reasoning were also examined. No reliable differences were found between scenarios (ps > .1) for references of the time frame of the change, how the change occurred, or the survival value of the change. Taken together, results support the conclusion that students preferentially applied intuitive reasoning as hypothesized—specifically, justifications using inheritance in the physical change scenario were more frequent than in the behavioral change scenario. The discovery of a difference in intuitive reasoning despite no reliable difference in scientific reasoning suggests that science-based and intuitive justifications were independently applied.

RQ#2 *Did students who learned about inheritance in an evolutionary context via category construction demonstrate more scientific reasoning and less intuitive reasoning than a worksheet control?*

Next, we examined differences between scientific and intuitive reasoning between category construction and the worksheet control. The difference in mean frequency of scientific justifications was examined between the category construction and worksheet control conditions using a Wilcoxon 2-Sample test bootstrapped with 1,000 iterations. Results showed that students assigned to category construction provided more scientifically accepted responses ($M = 1.12$, 95% CI = 0.95–1.28) than the worksheet control ($M = 0.47$, 95% CI = 0.29–0.68) in 98.7% of iterations of the analysis, $p = .013$ (see Figure 4). The mean frequency of intuitive justifications was also examined between the category construction and worksheet control conditions using a Wilcoxon 2-Sample test bootstrapped with 1,000 iterations. The analysis uncovered a significant difference between the conditions where category construction produced fewer intuitive explanations ($M = 0.67$, 95% CI = 0.55–0.79) than the worksheet control ($M = 1.13$, 95% CI = 0.90–1.37) in 95.5% of iterations of the analysis, $p = .045$ (see Figure 2). Taken together, these results suggest that category construction reduced the use intuitive

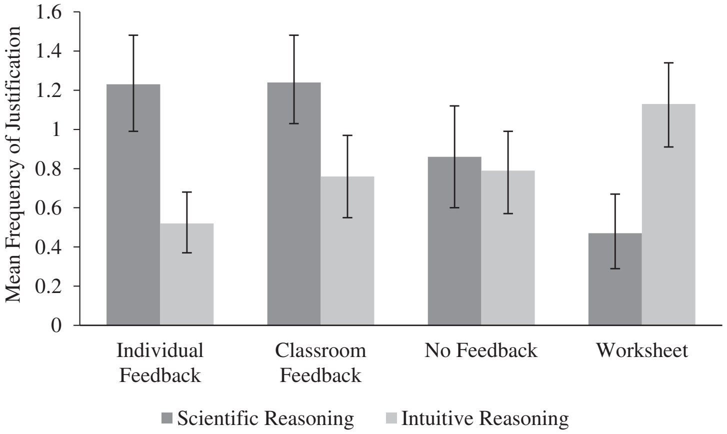


FIGURE 5 Comparison of the mean number of scientific and intuitive justifications for each condition for both scenarios combined. Error bars represent the 95% confidence interval

knowledge and increased the use of science-based knowledge when reasoning about heredity and evolutionary change.

RQ#3 *To what extent does feedback support student reasoning following category construction?*

Lastly, a secondary analysis was conducted to examine the relative contribution of different category construction conditions (classroom feedback, individual feedback, and no feedback) to the differences in student scientific and intuitive reasoning (see Figure 5). While there were no significant differences among category construction conditions, variations in implementation could have impacted student reasoning differentially in comparison to the worksheet control. Results showed that students who did not experience feedback following category construction ($M = 0.86$, 95% CI = 0.60–1.12) only demonstrated marginal increases in their use of science reasoning ($p = .07$) relative to worksheet students ($M = 0.47$, 95% CI = 0.29–0.67, $d = 0.40$). Students who experienced classroom feedback demonstrated significantly higher science reasoning ($p < .05$, $M = 1.24$, 95% CI = 1.03–1.48, $d = 0.93$) in comparison to worksheet. Similarly, students who experienced individual feedback also demonstrated higher science reasoning ($p < .05$, $M = 1.23$, 95% CI = 0.99–1.48, $d = 0.72$). Overall scientific reasoning was strongest when feedback was provided with category construction, but both levels (individual or classroom) appeared equally effective.

Next, differences in student intuitive reasoning were examined. Students who did not experience any feedback following category construction still demonstrated significant decreases in their use of intuitive reasoning ($p < .05$, $M = 0.79$, 95% CI = 0.55–0.97, $d = 0.37$) in comparison to the worksheet ($M = 1.13$, 95% CI = 0.91–1.34). Students who experienced classroom feedback also demonstrated significantly lower intuitive reasoning ($p < .05$, $M = 0.76$, 95% CI = 0.55–0.97, $d = 0.44$). Students who experienced individual feedback in the category construction task demonstrated the lowest intuitive reasoning ($p < .05$, $M = 0.52$, 95% CI = 0.37–0.68, $d = 0.76$) relative to worksheet. In total, implementation of the category construction task with classroom feedback produced the highest frequency of scientific reasoning. This was not the case for intuitive reasoning, where no feedback and classroom feedback showed similar effects. Intuitive reasoning was lowest when individual feedback was provided.

8 | DISCUSSION

The results support two conclusions: (1) Students can simultaneously hold and selectively apply science-based reasoning and intuitive understanding to explain inheritance based on a physical change cue; and (2) the category construction task appears to impact which of these reasoning strategies are applied. The category construction group applied more scientific explanations and fewer intuitive explanations for each scenario. While student accuracy during the categorization task is not addressed in this analysis, participating in the category construction learning activity is enough to improve the frequency of science-based explanations of the mechanisms of evolution and inhibit intuitive explanations. In addition, providing students with individual feedback appears to be the most optimal feedback method for the category construction task—in the present sample, it increased scientific reasoning and decreased intuitive reasoning to a greater degree than category construction without feedback. The task is lightweight, student-focused, and quite short in its duration, making the inclusion of feedback a reasonable addition to optimize student learning in the task.

Two implications arise from the coexistence and contextual nature of student reasoning observed in the current study. The first is that assessments of students must be cognizant of the contextual properties of misconceptions. A study including only Scenario 1 would have produced the conclusion that students did not believe that intergenerational change can impact inheritance. Scenario 2, however, clearly showed contextual application of this intuitive conception in the study population. The majority (61%) of students' reasoning for explicit inheritance in Scenario 2 argued against impact in Scenario 1, with only 2 students (0.9%) having reasoned for explicit inheritance in both scenarios. Extrapolating this finding to the population, over 60% of students might be expected to make the intuitive assertion that physical changes will have inheritance impacts. The results suggest that during initial exposure to evolution concepts, students are extremely susceptible to viewing evolution in the context of physical change. As heredity is one of the major components of both evolution by means of natural selection and genetics, instructors in both areas need to be aware of this intuition prior to being able to actively work against it in the classroom.

Second, the results align with a coexistence framework that argues that the significant alteration or removal of misconceptions (specifically those of intuitive origin) may not be a realistic end goal for science education. Here, the data were collected at the end of an evolution unit where students had already received significant exposure to natural selection. Yet, the results show that only a minority of students (41% and 33% respectively per scenario) were able to provide a scientifically accepted explanation. In contrast, students drew on significantly more intuitive reasoning in Scenario 2, suggesting that an inverse relationship between scientific and intuitive reasoning may be overly simplistic. Simply stated, more work needs to explore the nuanced relationship between human intuitive and scientific reasoning. If we cannot remove or change intuitive reasoning, then supporting student ability to inhibit intuitive responses could be a fruitful line of future research.

8.1 | Limitations

There are limitations to the current study that should be more comprehensively addressed in future work. First, the item contexts hold a substantial influence on student reasoning patterns. The current study only reports on differential student responses to two different free response scenarios. These scenarios were not meant to be exhaustive nor represent the "core" of student understanding of inheritance. Rather, they offered a case study for examining coexistence and contextual student application of ideas. Future work that builds upon this by using more diverse intuitive misconceptions could be influential in revealing the extent to which contexts drive intuitive versus scientific student reasoning. Second, the current study did not account for the impact of classroom learning on this contextual application. While significant effects were seen based on the category construction task, the task only occupied a minority of the time students spent learning about natural selection. As such, future work that tracks the impact of instructional variables on contextual student reasoning may be fundamental for understanding how the inhibition of intuitive reasoning occurs. Third, while the current study shows potential for the category construction strategy, its novelty in the classroom means that many unknowns remain at this point. One particular unknown of the current study that needs to be addressed in future work is the effectiveness of category construction in comparison to a worksheet with both examples and nonexamples. We choose to design the comparison worksheet to mirror the common practice of the teachers, but this meant that there was not a 1:1 information match between conditions. While contrast examples might be out of place in a worksheet activity designed to introduce a new concept, they are critical for the category construction task. It is clear that a better understanding of their role and design is needed to fully characterize the impact of learning via category construction. Fourth, the role of feedback during category construction needs to be more systematically compared to other tasks. The process by which students complete the category construction task is more complex than finding and copying specific pieces of information (the common "reading comprehension" practice for worksheets). This means that the worksheet and category construction tasks have differential built in support (beyond external feedback) for students. Future work needs to better characterize optimal implementation of the task, as well as how it compares to a diversity of other learning tools.

8.2 | Implications for practice

Beyond recognizing the importance of coexistence between scientific and intuitive conceptions, the present results suggest that category construction can increase the use of scientific reasoning over intuitive reasoning. Some instructors may wish to make use of category construction in the classroom. For these instructors, we would like to emphasize that to-be-sorted scenarios should be diverse in their superficial features and identical in their relational structure, that is, the way in which the scenarios exemplify the target principle. Feature-based sorting takes place when individuals use the surface attributes (similar colors, shapes, animals, habitats) of the scenarios to build groups. This might lead to the erroneous belief that certain animals or habitats adhere to the principle but others do not. Only the relational structure underlying the examples determines membership in the principle-based relational category. Therefore, care must be taken to create scenarios with a diversity of superficial features so that the only commonality is the way that the scenarios exemplify the target principle.

Examples of evolution and its constituent components (e.g., heredity, natural selection) can be better represented as members of a relational category (in comparison to a feature based category) because the examples are defined by consistent commonalities in relational structure (Goldwater & Schalk, 2016). Catalysis, force, predation, and flow are all examples of scientific concepts that can be viewed as relational categories—example members of these categories exemplify the interaction of a set of entities and their relationships. A more general example of a relational category is "family". This type of relational category not only includes explicit biological relationships (e.g., mother that gave birth to her daughter), but also hierarchical based social responsibilities (e.g., the mother takes care of her daughter) and expectations (e.g., older generations are obeyed by younger generations). Evolution can be thought of as complex relational category. Natural selection is just one of several mechanisms of evolution, but by itself includes many different components (e.g., phenotypic variation, limited resources, predation, competition, heritability, and mating) with typical relational roles interacting dynamically within a particular environment. For example, population limiting forces (e.g., disease, predation, food, and shelter) impact individual's survival based on their phenotypic variation. Surviving individuals are more likely to reproduce. If the phenotypic variation has a genetic component, then it has a chance to be inherited by offspring. This differential survival paired with heritability results in shifts in gene frequencies within a population's gene pool (i.e., evolutionary change). The specific way in which these relationships play out can vary, but the broad relational structure laid out above is consistent for any situation that can be classified as an instance of natural selection.

The process of knowledge development in science often emerges from comparison and abstraction among multiple sources of data. Category construction can be seen as representing this process of knowledge development as it requires students to compare across exemplars and then abstract a larger idea from them. The target of category construction can take many forms. In some instances the target idea may be the scientific theory itself and in other instances the target may be smaller components (we refer to this as a principle) of a theory. It is not uncommon for evolutionary theory to be approached via its component parts (for a review see Gregory, 2009) so in this study we chose to target just one aspect of this larger theory due to well-documented difficulty that students have in mastering evolutionary theory (Bishop & Anderson, 1990; Demastes et al., 1995; Gregory, 2009; Smith, 2010). This does not mean that theories could not be targeted using category construction in future work.

Category construction offers an opportunity for students to look across multiple scenarios to abstract the relationships that define evolutionary change for themselves. In this study, the aim was to promote the development of an abstraction that underlies the distinctions between genetic and phenotypic change for offspring characteristics. We can imagine more intense uses of category construction in the classroom where the principle is not presented ahead of time to students—though in this study's context the provision of the principle seemed necessary. Even without a principle being presented a head of time, students would still be provided with a number of varied scenarios that encourage them (with instructor feedback) to abstract relationships between different components of natural selection. This student-driven discovery approach may have better results than a direct instruction approach with more advanced students that can be relied on to extract the relational commonalities on their own, as is found with college-level learners (Kurtz & Honke, under review).

8.3 | Conclusion

The current study emphasized the value of intuitive student conceptions beyond the obstructive characterization they often receive within the science learning community. Under this perspective, student reasoning about inheritance was examined under contextual variation and the results revealed that intuitive and science-based reasoning coexists in middle level science student cognition. Further, the category construction task seeded with scientifically valid scenarios can promote application of science-based reasoning over intuitive reasoning. These findings are not limited to the science learning context, but also raise questions about how intuitive conceptions/misconceptions should be approached in the classroom and the potential limitations of assessments (which are often not designed to address contextual variation in student intuitive responses). These assessment implications can be particularly slippery for educators. At what point can an assessment be considered contextually sensitive enough to provide robust measures of student science understanding? One useful step will be to identify the contexts most likely to elicit intuitive student conceptions—a fruitful place to begin may be contexts in which an intuitive conception could be beneficial to everyday human functioning.

ACKNOWLEDGMENTS

Work funded by a grant from IES Cognition and Student Learning #R305A120554. The authors would like to thank Nolan Conway, John D. Patterson, and Sean Snoddy for their contributions. The authors declare no conflicts of interest relative to this manuscript.

ENDNOTE

[1] The authors recognize that a myriad of terms have been used to (re)classify student ideas that run counter to science understandings (Leonard, Kalinowski, & Andrews, 2014) and have chosen intuitive conceptions to discuss a subset of these which have everyday value. This can be synonymous with the term misconception in some cases.

ORCID

Joshua Premo http://orcid.org/0000-0001-6739-4479

REFERENCES

Abraham, J. K., Perez, K. E., Downey, N., Herron, J. C., & Meir, E. (2012). Short lesson plan associated with increased acceptance of evolutionary theory and potential change in three alternate conceptions of macroevolution in undergraduate students. *CBE-Life Sciences Education*, *11*(2), 152–164.

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist*, *48*(2), 87–113.

Alters, B. J., & Nelson, C. E. (2002). Perspective: Teaching evolution in higher education. *Evolution*, *56*(10), 1891–1901.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612.

Bishop, B. A., & Anderson, C. W. (1990). Student conceptions of natural selection and its role in evolution. *Journal of Research in Science Teaching*, *27*(5), 415–427.

Bryce, T. G. K., & MacMillan, K. (2009). Momentum and kinetic energy: Confusable concepts in secondary school physics. *Journal of Research in Science Teaching*, *46*(7), 739–761.

Cavagnetto, A. R., & Kurtz, K. J. (2016). Promoting students' attention to argumentative reasoning patterns. *Science Education*, *100*(4), 625–644.

Chen, Y. L., Pan, P. R., Sung, Y. T., & Chang, K. E. (2013). Correcting misconceptions on electronics: Effects of a simulation-based learning environment backed by a conceptual change model. *Educational Technology & Society*, *16*(2), 212–227.

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, *70*(5), 1098–1120.

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152.

Clark, D. B. (2006). Longitudinal conceptual change in students' understanding of thermal equilibrium: An examination of the process of conceptual restructuring. *Cognition and Instruction*, *24*(4), 467–563.

Coley, J. D., & Tanner, K. D. (2012). Common origins of diverse misconceptions: Cognitive principles and the development of biology thinking. *CBE-Life Sciences Education*, *11*(3), 209–215.

de Posada, D., & María, J. (1997). Conceptions of high school students concerning the internal structure of metals and their electric conduction: Structure and evolution. *Science Education*, *81*(4), 445–467.

Demastes, S. S., Settlage, J., & Good, R. (1995). Students' conceptions of natural selection and its role in evolution: Cases of replication and comparison. *Journal of Research in Science Teaching*, *32*(5), 535–550.

Dennett, D. C. (1989). *The intentional stance*. Cambridge, MA: MIT Press.

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

Emmons, N. A., & Kelemen, D. A. (2015). Young children's acceptance of within-species variation: Implications for essentialism and teaching evolution. *Journal of Experimental Child Psychology*, *139*, 148–160.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*(1), 255–278.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241.

Field, A. (2013). Discovering statistics using SPSS. London: SAGE Publications.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York, NY: Oxford University Press.

Gelman, S. A. (2011). When worlds collide–or do they? Implications of explanatory coexistence for conceptual development and change. *Human Development*, *54*(3), 185–190.

Gelman, S. A., & Rhodes, M. (2012). Two-thousand years of stasis. How psychological essentialism impedes evolutionary understanding. In K. S. Rosengren, S. Brem, E. M. Evans, & G. Sinatra (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 3–21). New York, NY: Oxford University Press.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*(2), 155–170.

Gentner, D., & Kurtz, K. (2005). Relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. W. Wolff (Eds.), *Categorization inside and outside the lab* (pp. 151–175). Washington, DC: APA.

Gentner, D., Loewenstein, J., Thompson, L., & Forbus, K. (2009). Reviving inert knowledge: Analogical abstraction supports relational retrieval of past events. *Cognitive Science*, *33*(8), 1343–1382.

Gentner, D., Rattermann, M. J., & Forbus, K. D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology*, *25*(4), 524–575.

Gericke, N. M., & Hagberg, M. (2007). Definition of historical models of gene function and their relation to students' understanding of genetics. *Science & Education*, *16*(7–8), 849–881.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*(3), 306–355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), 1–38.

Glynn, S. M. (1991). Explaining science concepts: A teaching-with-analogies model. In S. Glynn, R. Yeany, & B. Britton (Eds.), The psychology of learning science (pp. 219–240). Hillsdale, NJ: Erlbaum.

Goldwater, M. B., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research. *Psychological Bulletin*, *142*(7), 729.

Gregory, T. R. (2009). Understanding natural selection: essential concepts and common misconceptions. *Evolution: Education and Outreach*, *2*(2), 156–175.

Gregory, T. R., & Ellis, C. A. (2009). Conceptions of evolution among science graduate students. *BioScience*, *59* (9), 792–799.

Ha, M., Lee, J. K., & Cha, H. Y. (2006). A cross-sectional study of students' conceptions on evolution and characteristics of conception formation about it in terms of the subjects: Human, animals and plants. *Journal of Korean Association for Research in Science Education*, *26*(7), 813–825.

Haslam, N., & Whelan, J. (2008). Human natures: Psychological essentialism in thinking about differences between people. *Social and Personality Psychology Compass*, *2*(3), 1297–1312.

Heddy, B. C., & Sinatra, G. M. (2013). Transforming misconceptions: Using transformative experience to promote positive affect and conceptual change in students learning about biological evolution. *Science Education*, *97*(5), 723–744.

Honke, G., Cavagnetto, A. R., Kurtz, K. J., Patterson, J. D., Conaway, N. B., Tao, Y., & Marr, J. C. (2015). *Promoting transfer and mastery of evolution concepts with category construction*. In Paper presented at the American Educational Research Association annual meeting, Chicago, IL.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Macmillan.

Kaminski, G., Gentaz, E., & Mazens, K. (2012). Development of children's ability to detect kinship through facial resemblance. *Animal Cognition*, *15*(3), 421–427.

Kampourakis, K., Palaiokrassa, E., Papadopoulou, M., Pavlidi, V., & Argyropoulou, M. (2012). Children's Intuitive teleology: Shifting the focus of evolution education research. *Evolution: Education and Outreach*, *5*(2), 279–291.

Kelemen, D. (1999). The scope of teleological thinking in preschool children. *Cognition*, *70*(3), 241–272.

Kelemen, D. (2012). Teleological minds: How natural intuitions about agency and purpose influence learning about evolution. *Evolution Challenges: Integrating Research and Practice in Teaching and Learning about Evolution*, *66*, 92.

Klahr, D., & Li, J. (2005). Cognitive research and elementary science instruction: From the laboratory, to the classroom, and back. *Journal of Science Education and Technology*, *14*(2), 217–238.

Kurtz, K. J., Boukrina, O., & Gentner, D. (2013). Comparison promotes learning and transfer of relational categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1303–1310.

Kurtz, K. J., Cavagnetto, A. R., Honke, G., Conaway, N. B., Patterson, J. D., Marr, J. C., & Tao, Y. (2014). Optimizing the category construction task to promote learning and transfer of knowledge in classroom instruction. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Kurtz, K. J., & Honke, G. (under review). *Sorting out the problem of inert knowledge: Category construction to promote spontaneous transfer*.

Kurtz, K. J., Miao, C. H., & Gentner, D. (2001). Learning by analogical bootstrapping. *The Journal of the Learning Sciences*, *10*(4), 417–446.

Lamb, R., & Premo, J. (2015). Computational modeling of teaching and learning through application of evolutionary algorithms. *Computation*, *3*(3), 427–443.

Legare, C. H., Lane, J. D., & Evans, E. M. (2013). Anthropomorphizing science: How does it affect the development of evolutionary concepts? *Merrill-Palmer Quarterly*, *59*(2), 168–197.

Legare, C. H., & Visala, A. (2011). Between religion and science: Integrating psychological and philosophical accounts of explanatory coexistence. *Human Development*, *54*(3), 169–184.

Leonard, M. J., Kalinowski, S. T., & Andrews, T. C. (2014). Misconceptions yesterday, today, and tomorrow. *CBE-Life Sciences Education*, *13*(2), 179–186.

Lewis, J., & Kattmann, U. (2004). Traits, genes, particles and information: re-visiting students' understandings of genetics. *International Journal of Science Education*, *26*(2), 195–206.

Loewenstein, J. (2010). How one's hook is baited matters for catching an analogy. *Psychology of Learning and Motivation*, *53*, 149–182.

Loewenstein, J., Thompson, L., & Gentner, D. (1999). Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin & Review*, *6*(4), 586–597.

Marbach-Ad, G. (2001). Attempting to break the code in student comprehension of genetic concepts. *Journal of Biological Education*, *35*(4), 183–189.

Marbach-Ad, G., & Stavy, R. (2000). Students' cellular and molecular explanations of genetic phenomena. *Journal of Biological Education*, *34*(4), 200–205.

Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory & Cognition*, *24*(2), 235–249.

Markman, A. B., & Gentner, D. (1997). The effects of alignability on memory storage. *Psychological Science*, *8*(5), 363–367.

Matlen, B. J., & Klahr, D. (2013). Sequential effects of high and low instructional guidance on children's acquisition of experimentation skills: Is it all in the timing? *Instructional Science*, *41*(3), 621–634.

Miller, J. D., Scott, E. C., & Okamoto, S. (2006). Science communication: Public acceptance of evolution. *Science*, *313*(5788), 765–766.

Mortimer, E. F. (1995). Conceptual change or conceptual profile change? *Science & Education*, *4*(3), 267–285.

Nehm, R. H., Beggrow, E. P., Opfer, J. E., & Ha, M. (2012). Reasoning about natural selection: diagnosing contextual competency using the ACORNS instrument. *The American Biology Teacher*, *74*(2), 92–98.

Nehm, R. H., & Ha, M. (2011). Item feature effects in evolution assessment. *Journal of Research in Science Teaching*, *48*(3), 237–256.

Oliver, M. (2011). Teaching and learning evolution: Testing the principles of a constructivist approach through action research. *Teaching Science: The Journal of the Australian Science Teachers Association*, *57*(1), 13–18.

Pugh, K. J., Koskey, K. L., & Linnenbrink-Garcia, L. (2014). High school biology students' transfer of the concept of natural selection: a mixed-methods approach. *Journal of Biological Education*, *48*(1), 23–33.

Sabella, M. S., & Redish, E. F. (2007). Knowledge organization and activation in physics problem solving. *American Journal of Physics*, *75*(11), 1017–1029.

Schneider, M., & Hardy, I. (2013). Profiles of inconsistent knowledge in children's pathways of conceptual change. *Developmental Psychology*, *49*(9), 1639.

Shtulman, A., & Calabi, P. (2012). Cognitive constraints on the understanding and acceptance of evolution. In K. S. Rosengren, S. K. Brem, E. M. Evans, & G. M. Sinatra (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 47–65). New York: Oxford University Press.

Shtulman, A., & Lombrozo, T. (2016). Bundles of contradiction: A coexistence view of conceptual change. In D. Barner & A. Baron (Eds.), *Core knowledge and conceptual change*. Oxford, UK: Oxford University Press.

Shtulman, A., Neal, C., & Lindquist, G. (2016). Children's ability to learn evolutionary explanations for biological adaptation. *Early Education and Development*, 1–15.

Shtulman, A., & Schulz, L. (2008). The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science*, *32*(6), 1049–1062.

Sinatra, G. M., Brem, S. K., & Evans, E. M. (2008). Changing minds? Implications of conceptual change for teaching and learning about biological evolution. *Evolution: Education and Outreach*, *1*(2), 189–195.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3.

Smith, M. U. (2010). Current status of research in teaching and learning evolution: II. Pedagogical issues. *Science & Education*, *19*(6–8), 539–571.

Tamir, P., & Zohar, A. (1991). Anthropomorphism and teleology in reasoning about biological phenomena. *Science Education*, *75*(1), 57–67.

Thorn, C. J., Bissinger, K., Thorn, S., & Bogner, F. X. (2016). Trees live on soil and sunshine!"-coexistence of scientific and alternative conception of tree assimilation. *PLoS One*, *11*(1), e0147802.

Venville, G., Gribble, S. J., & Donovan, J. (2005). An exploration of young children's understandings of genetics concepts from ontological and epistemological perspectives. *Science Education*, *89*(4), 614–633.

Venville, G. J., & Treagust, D. F. (1998). Exploring conceptual change in genetics using a multidimensional interpretive framework. *Journal of Research in Science Teaching*, *35*(9), 1031–1055.

Vosniadou, S. (1992). Knowledge acquisition and conceptual change. *Applied Psychology*, *41*(4), 347–357.

Williams, J. J., & Lombrozo, T. (2010). The role of explanation in discovery and generalization: Evidence from category learning. *Cognitive Science*, *34*(5), 776–806.

Williams, J. J., & Lombrozo, T. (2013). Explanation and prior knowledge interact to guide learning. *Cognitive Psychology*, *66*(1), 55–84.

Williams, J. M. (2012). Children and adolescents' understandings of family resemblance: A study of naïve inheritance concepts. *British Journal of Developmental Psychology*, *30*(2), 225–252.

Williams, J. M., & Smith, L. A. (2010). Concepts of kinship relations and inheritance in childhood and adolescence. *British Journal of Developmental Psychology*, *28*(Pt 3), 523–546.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Premo J, Cavagnetto A, Honke G, Kurtz KJ. Categories in conflict: Combating the application of an intuitive conception of inheritance with category construction. *J Res Sci Teach*. 2018;00:1–21. https://doi.org/10.1002/tea.21466