Please cite as:

Asterhan, C. S. C. & Schwarz, B. B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. *Journal of Educational Psychology*, *99*, 626-639

The Effects of Monological and Dialogical Argumentation on Concept Learning in Evolutionary Theory

Christa S. C. Asterhan and Baruch B. Schwarz

School of Education

The Hebrew University of Jerusalem

Mt Scopus, Jerusalem 91905, Israel

Email : asterhan@huji.ac.il , msschwar@mscc.huji.ac.il

Abstract

In this study the effects of argumentation-eliciting interventions on conceptual understanding in evolution were investigated. Two experiments were conducted: In the first, 76 undergraduates were randomly assigned to dyads to collaboratively solve and answer items in evolution; half of them were instructed to conduct an argumentative discussion, whereas control dyads were only asked to collaborate. In the second experiment, 42 singletons participated in one of two conditions: Experimental students engaged in monological argumentation on their own and a confederate's solution in response to prompts read by the confederate, whereas in the control condition they merely shared their solutions. Conceptual gains were assessed on immediate and delayed post-tests. In both experiments, students in the argumentative conditions showed larger learning gains on the delayed post-test than control students. Students in argumentative conditions were able to preserve gains that were obtained immediately following the intervention, whereas control subjects either lost immediate gains (dialogical condition) or did not improve their conceptual understanding at any time (monological condition).

Keywords: Argumentation, concept learning, collaborative learning, evolutionary theory, cognitive conflict.

The effects of Monological and Dialogical Argumentation on Concept Learning in Evolutionary Theory

Many practitioners and researchers have identified argumentative practices in the classroom as an important means of improving students' understanding of subject matter (e.g., Baker, 2003; De Vries, Lund & Baker, 2002; Duschl & Osborne, 2002; Mason, 2001; Nussbaum & Sinatra, 2003; Schwarz, 2003; Schwarz & Glassner, 2003; Schwarz, Neuman & Biezuner, 2000). Such practices are best defined as activities in which interlocutors cooperate in solving a particular problem to which a number of different solutions are proposed. Each of these solutions is perceived to be of a different epistemic status and interlocutors feel obliged to choose between them (Baker, 1999; 2003). In contrast to other forms of conflict resolution, they attempt to reach this goal by engaging in reasoning.

However, these potentially beneficial effects have thus far not been successfully subjected to rigorous experimental testing. In this paper, we present results from two experimental studies on the effects of argumentation on conceptual understanding of evolutionary theory. Specifically, we propose that the engagement in argumentation by producing reasoned arguments in favour of and against one’s own and another person’s views will facilitate concept learning. The learning tasks in these two studies were designed within the cognitive conflict paradigm. Within this instructional paradigm, students’ naïve views are either confronted with anomalous data or contradicting views or are paired with peers who have different views (Limon, 2001).

The expectation that argumentation will foster concept learning within these settings is based on the fact that argumentation intertwines a number of social and cognitive processes considered to promote concept learning:

First of all, similar to the self-explanation effect, the examination of one’s theories and the reasons behind them has been known to promote understanding and knowledge construction processes (e.g., Chi, 2000; Chi, Bassok, Lewis, Reimann & Glaser, 1989; Chi, de Leeuw, Chiu & Lavancher, 1994; Kuhn, 1991; Neuman & Schwarz, 2000). Argument generation in social settings may further encourage clarification of contradictions in one's conceptual understanding, especially when one is attempting to convince others. Research on accountability effects has shown that even the mere anticipation of an unknown audience has been found to improve the quality of thinking (e.g., Tetlock, 1992).

Furthermore, in *dialectical* argumentation participants are exposed to a multiplicity of ideas and encouraged to explore the validity of each other's ideas. This implies that they have to consider objections to their theories and assumptions, attempt to understand alternative positions and formulate objections and counter-objections. In addition, the unique structure of argumentation that links premises, conclusions, conditions, rebuttals and so forth is thought to considerably improve the organization of knowledge (Billig, 1996; Kuhn, 1992; Means & Voss, 1996). In turn, this should arguably lead to better recall and understanding on subsequent test occasions

In theory, engaging in solitary argumentation should yield similar results. However, argumentation is basically a social process that presupposes the presence of an audience (Leitao, 2000). Moreover, the act of solitary argumentation about scientific concepts is likely to be cognitively too demanding. The presence of a dialogic partner, on the other hand, may promote reflection and awareness of one's own beliefs (Amigues, 1988), cause learners to engage in explanatory activities (Okada & Simon, 1997) and reduce cognitive load through the personification of different alternatives (Baker, 2003).

Some experimental studies have been conducted on the effects of argumentative *writing* on learning and various performance measures. For example, in a study on the effects of different writing instructions, Wiley & Voss (1999) found that instructions to write an argumentative essay on an historical topic produced better conceptual understanding than other instructions (narrative, summary or explanation), but not better retention of factual information.

A notable exception in the argumentation and learning literature is a study conducted by Nussbaum & Sinatra (2003), who investigated the effects of a conceptual change intervention based on argumentation. In this study, undergraduates were asked to predict the path of a falling object dropped in different settings and subsequently provide explanations for their choices. After each prediction, subjects in the experimental condition were instructed to argue in favor of a given alternative prediction (i.e., the correct solution). Participants in both experimental and control conditions were shown a simulation of the correct solution before continuing to the next item. The two conditions did not differ in the accuracy of their predictions, but the explanations in the experimental condition were of a higher quality than in the control condition.

However, their experimental intervention also included prompts that were meant to help subjects formulate reasons. The administration of these prompts was based on and tailored to whether the subject had shown recognition of key aspects of conceptual understanding in counterarguments. As a result, it is not entirely clear whether the difference in the quality of the explanations in this study was due to the tailored prompts or the fact that subjects were asked to provide counterarguments.

The main purpose of the two studies presented here was to investigate the effect of oral argumentation on conceptual understanding of a scientific topic, in a controlled experimental design. Specifically, we tested whether interventions that elicit argumentation would yield larger gains in conceptual understanding than control conditions. We present two experiments: In the first, we examined whether collaborative learning tasks involving dialogical argumentation would yield better conceptual gains than ordinary collaborative learning tasks. In the second, individuals engaged in monological argumentation prompted by a confederate.

The scientific concept we chose to focus on is evolutionary theory. We had several reasons for this choice: First of all, evolution may be particularly appropriate for a study on argumentation in isolation, since teaching evolution is mostly about choosing the most plausible explanation (Ohlsson, 1992). Other scientific concepts -such as *mass* or *light*- may benefit more from instructional methods that involve hypothesis testing strategies (e.g., Howe, Tolmie, Duchak-Tanner & Rattay, 2000). Secondly, even without formal education on the topic, most people have intuitive theories about evolution. Research has, furthermore, shown that traditional teaching methods have overall failed in this field, whereas some findings indicate that collaborative peer activities involving reasoning about one another's ideas may be the key to effective learning on this topic (Jensen & Finley, 1996; Jimenez-Aleixandre, 1992). Moreover, this concept has been extensively studied over the years and several researchers have elaborated criteria for defining knowledge about this concept (e.g., Bishop & Anderson, 1990; Brumby, 1984; Greene, 1990; Ohlsson & Bee, 1992).

An additional reason is that evolution is a complex concept whose correct understanding requires the integration of a number of different principles. This allows for assessment on two different levels: the schemata students use to explain evolutionary phenomena and the number of correct principles they produce in their explanations. Even though the former have generally been referred to as *mental models* in the concept learning literature (e.g., Chi & Roscoe, 2003; Vosniadou & Brewer, 1994), we prefer the term *explanatory schemas* (Ohlsson, 2002). While explanations are content-specific generative descriptions, explanatory schemas encode the structure shared by a set of explanations. This structure is defined by the set of generative relations and the way they are combined (Ohlsson, 2002).

This view agrees with current theoretical frameworks in the concept learning literature (e.g., Samarapungavan, & Wiers, 1997; Schnotz & Preuss, 1999), according to which individuals have a repertoire of schemas for a specific type of event and each schema configuration has a different likelihood of being activated according to its previous usefulness in specific contexts. Hence, an individual may use different schemas to explain different evolutionary events.

Research has shown that most teenagers do not possess adequate argumentative skills, and that higher education is a significant predictor of differences among adults in most aspects of such skills (Kuhn, 1991). To avoid the inclusion of an intensive instructional unit on argumentation, the two studies presented here were conducted on undergraduates. Clearly, undergraduates are a more selective sample: Most of them are expected to be capable of conducting an argumentative discussion, while not being too experienced in critically discussing scientific topics.

Study 1: Dialogical Argumentation

In the first study we presented students with the scientifically accepted account of evolutionary change, after which they were assigned to dyads and solved two questions on evolutionary phenomena. Some of these dyads were instructed to engage in argumentative discussion, whereas others were merely instructed to solve the items collaboratively. Our hypotheses are as follows:

Conceptual change tasks in which collaborators are instructed to engage in argumentation should yield larger learning gains, than situations in which they are not instructed to do so. In addition, we expect that some control dyads may naturally engage in dialectical argumentation without being specifically instructed to do so, whereas some experimental dyads may not be able or willing to comply with the instructions. Learning gains should be dependent on such naturally occurring within-group differences.

Method

Participants

Eighty-six undergraduates (freshmen and sophomores) in Education (63%), Social Sciences (9%) and Humanities (28%) at the Hebrew University participated in this study (mean age = 22.8). They were offered the choice between course credit (71%) or financial reward (29%) for their participation. Students were recruited through announcements on note-boards around campus. Ten participants either had difficulties with Hebrew or showed minimum effort in the pretest session (i.e., finished the questionnaire in less than 6 minutes), or were paired with such a participant in the intervention phase of the experiment. Their data was omitted from the analyses. The final sample included students with different backgrounds in formal biology education: 19% reported they had been taught the topic of evolution, 78% reported they had not, and two Ss failed to report.

Design

A pre-test, intervention, and post-test experimental design was used. The two conditions differed in the amount of support and encouragement dyads were given to engage in a dialectical argumentative discussion (see Procedure section). Individual conceptual understanding in evolution was assessed on three separate test occasions (pretest, immediate post-test and delayed post-test). A preliminary pilot study was conducted to determine the effectiveness and the exact features of the manipulation and the format and content of the test items in this study. The results of this pilot study will not be discussed in detail, but will be mentioned when relevant.

Procedure

All subjects participated in the following sequence of activities: (1) pre-test (T1) to assess prior evolutionary understanding; (2) instructional intervention, which was identical in all conditions (screening of instructional movie excerpt); (3) experimental intervention in which dyads solved two items on evolutionary phenomena according to two different conditions (control and argumentative); (4) immediate post-test (T2), which assessed evolutionary understanding after the intervention; (5) delayed post-test (T3), administered a week later. The total length of each session was approximately one and a half hours. Tests at T1 and T3 were administered to individuals in a group format (size ranging from 2 to 8) as paper-and-pencil tests.

Pretest (T1). At the start of the experiment subjects were informed that the experiment dealt with people's understanding of evolutionary processes, that their participation was anonymous and that the two best performances would each win a financial prize (coupons of approximately $23, valid in campus stores). They then filled out a short demographic survey in which, among others, they reported on formal education in evolution. Following, they were administered the pretest. All participants finished the questionnaires within the 25-minute time limit.

Instructional intervention. Following T1, the educational movie excerpt was screened. Subjects were requested to pay close attention, refrain from interaction with each other and take notes if they wanted to. They were also told that the movie presented the Darwinian explanation of evolutionary processes and that they would be asked about evolution in the next two phases of the experiment. The movie excerpt showed that a number of modern Galapagos species were different from their continental ancestors (iguanas, turtles, sea-lions), without an account of how that change had occurred. In addition, changes in a population of Galapagos finches (also known as Darwin's finches) were discussed in detail and explained in terms of the Darwinian account of evolutionary change. This explanation also included a step-by-step graphical presentation. Although this detailed explanation referred to five of the six Darwinian principles that were assessed in this study (see Coding procedures section), the source of intra-species variability was not discussed in the movie excerpt. Darwinism was presented as *the* scientifically proven and accepted theory of evolution today. No alternative explanations were mentioned.

Experimental intervention and immediate post-test (T2). After the movie, subjects were randomly assigned to a partner and each dyad was assigned to one of the two conditions. Each dyad was placed in either separate rooms or separate, shielded working corners within the same room, depending on the group size of the experimental session. Each participant received a four-page booklet (see Materials section). The experimenter read the instructions and prompts aloud to the participants and then left the room. The dyads then collaboratively solved two items on evolutionary change processes (one warm-up and one target item) without any time limit while being audio-taped. After discussing the solution to an item, subjects wrote the solution separately, each on their own personal answer sheet.

So as to secure an actual difference in argumentation between the conditions, the manipulation was based on results from a preliminary pilot study and existing findings on the conditions that foster productive argumentation (e.g., Kuhn & Lao, 1998, Kuhn, Felton & Shaw, 1997; Schwarz, Neuman, Gil & Ilya, 2003). Accordingly, the two conditions differed in two features:

(a) Content of instructions: Instructions in the control condition were as follows:

> In this part of the experiment you are requested to solve the following two questions <u>together</u>. Try to reach the best and most comprehensive solution to each question. Consider the proposed solution well before writing the answer on your sheet. It is very important to elaborate your answers by discussing each relevant concept and idea in as detailed and comprehensive a way as possible. Each participant will write the answer to each question on his/her own answer sheet.

The instructions in the experimental condition were as follows:

> In this part of the experiment you are requested to solve the following two questions by conducting a collaborative critical discussion. Try to reach the best and most comprehensive solution to each question. Each participant should write the answer to each question on his/her own answer sheet. Before doing so, you should conduct an in-depth discussion by arguing for and against each position, solution and idea that is proposed during the discussion and by providing adequate justifications for them. In a critical discussion each person proposes the solution he or she thought of. You will then try to persuade each other and explain why you think a certain solution is better than the other. The goal of this discussion is to reach the best possible solution together, through persuasion and by producing both supporting and refuting arguments. Following are a number of questions that may guide you in conducting a critical discussion:
>
> * Why is your claim true? Try to explain, elaborate and justify why your solution is true or better.
> * Why is a certain idea or solution incorrect? Elaborate on the reasons.
> * Can you provide evidence for your claim?
> * Can you prove the incorrectness or weakness of a certain argument or solution?
> * What are the weaknesses in your or your partner's arguments?
> * To what extent do the justification, the proof and the explanations really support the proposed claim?

Experimental subjects were then instructed to first think independently about the solution to the test item, before starting the discussion on their respective answers.

(b) Argumentative model: Experimental dyads also received a short excerpt of a critical discussion of four turns between two (hypothetical) subjects who, they were told, had participated in the experiment a year earlier. The excerpt was as follows:

> A: Then the ducks had to change their feet so that they could swim. The area was flooded with water, and because of the new environment webbed feet developed.
>
> B: What do you mean "developed"? How did that happen?
>
> A: Hmmmm. In the beginning they did not know how to swim. But slowly they learned to do it and that caused some sort of development in their feet. I mean, webs developed between their fingers. And that's how it was passed on to the next generation.
>
> B: Well, if that were true, then Olympic swimmers should also develop webbed feet, since they also swim all day long!

The dyads were given the excerpt while they had been working on the target question (the Duck question) for at least 30 seconds, so as to allow them to at least to articulate their own solutions to the Duck item. They were told that the experimenter ”forgot” to give them the item in time and they were asked to read it and consider it in their discussion. The discussion in the excerpt modeled a critical dialog on the Duck item without actually revealing or hinting at the correct solution. Results from the pilot study showed, that argumentation on a scientific topic has to be actively sustained in order to avoid premature closure of the argumentative discussion.

Delayed post-test. After one week Ss were administered the delayed post-test under conditions similar to the pretest procedure. They received financial rewards and/or course credit upon completion of the post-test.

Materials

(a) A 20-minute excerpt of an educational movie in the "New Frontiers" series, which presents Darwin's theory of the evolution of species, with Hebrew subtitles (see Procedure).

(b) Pre- and delayed post-test questionnaires each containing three test items: one warm-up and two target items.

(c) A four-paged booklet for use during the intervention, which included written instructions, two immediate post-test items (one warm-up and one target) and individual answer sheets.

Three slightly different types of open test items were developed, a warm-up and two different types of target items. The types had structural similarity but different surface features (see Table 1 for examples of each type). They were adapted from existing research on evolutionary understanding (Ferrari & Chi, 1998; Jensen & Finley, 1995; Sandoval, 2003):

In the warm-up item Ss were asked to predict how a population of animals (butterflies, finches) would change, given a certain environmental demand. In these items, intra-species variation was an explicit given of the data. Pilot study results showed that Darwinian accounts were therefore more easily elicited than in other formats.

The two types of target items both contained data on how a certain species' ancestors used to be different in a given trait. Subjects were then requested to explain how evolutionary theory would describe this process of change. The first type of target item referred to traits for which it is difficult to imagine how an individual might intentionally improve on them (i.e., peppered moths’ changing coloration). Moreover, in this type of questions the term "population" was used, which implicitly hinted at existing variation. The second type of target item also required Ss to explain how a certain species evolved over time. However, the question was formulated in terms of a species instead of a population (e.g., "the cheetah"). In addition, the trait that evolved was of the type that could easily elicit Lamarckian accounts (e.g., running and swimming), since Ss can easily imagine how an individual might improve on them.

Both the pretest and the delayed post-test questionnaires contained three test items, one of each type. The warm-up item was identical in both, whereas the target items used different species (see Table 2). The intervention and the following immediate post-test contained two items: (1) a warm-up question on Darwin’s finches (a phenomenon described in the movie); (2) a target item on the evolution of the webbed feet of ducks. This item was chosen based on pilot study findings indicating that it produced cognitive conflict in many Ss.

Coding procedures

Conceptual understanding. The dependent measure of conceptual understanding of evolutionary processes was assessed according to two complementary means of scoring, both of which are commonly found in the literature: (1) a classification of the explanatory schemata that underlie students' explanations of evolutionary phenomena on each test item, as well as (2) a more fine-grained coding system intended to assess the number of Darwinian principles students produced in each response. Even though the outcomes of these different means of assessment are highly correlated (i.e. ranging .80 to .85 in this study), their correlation is by no means perfect: Individuals may use a Darwinian schema of evolutionary change to explain a phenomenon with either many or a minimum of correct principles, and vice-versa. Both coding schemes were constructed in consultation with experts on research in evolutionary biology education and validated by them.

(a) Explanatory schema score: Based on previous works (i.e., Brumby, 1984; Bishop & Andersen, 1990; Greene, 1990) we identified ten qualitatively different explanatory schemas in students' explanations of evolutionary change. Each different schema was then assessed quantitatively on four different dimensions: whether evolutionary change was considered at all, whether this change was explained, whether some sort of selection mechanism was used and whether existing intra-species variation was considered. For example, a Lamarckian schema of evolution acknowledges the occurrence of change among species, but does not consider intra-species variability or selection. This schema explains change in terms of a gradual process of adaptation of the group as a whole: Changes in traits are acquired in response to specific needs and passed on to the next generation.

Based on the appearance of each one of these four dimensions, each one of the ten qualitatively different explanatory schemas was assigned to one of five different categories. For example, an explanatory schema we called "mutations after" attributes evolutionary change to mutations that occur in all individuals in a population in reaction to a change in the environment. These changes in genetic make-up are then passed on to offspring. Even though the mechanism described here is distinctively different from the Lamarckian account, they both acknowledge and explain change, whereas the explanation does not consider intra-species variation and selection. They are therefore assigned to the same category of typological change. The categories and their corresponding explanatory schemas are described briefly in Table 3.

The score for each schema category was based on the number of dimensions in a given response. An additional null category was added to distinguish between explanations that did not consider change (the lowest category) and those responses that simply did not answer the question at all (by stating that they did not know the answer or repeating the data given in the item without providing additional information). The assessment, thus, yielded a scoring scale ranging from 0 to 5 (see Table 4). Responses in which two different possible solutions were offered, each corresponding to a different explanatory schema, were assigned the average of the two respective scores.

(b) Darwinist principles: The complementary coding scheme for Darwinist principles is adapted from Ohlsson's summary of modern evolution theory in the form of five key principles (Ohlsson & Bee, 1992). We omitted the original principle of *heritability* (only genetic traits may be passed on), and subdivided the original principles of *random intra-species variability* and *accumulation of changes over many generations*, thus resulting in the six principles presented in Table 5.

The appearance of each principle in each item was assessed according to the following grading key: the principle is mentioned and used correctly (2 points), only part of the principle is mentioned and used correctly (1 point), or the principle is not mentioned at all (0 points). The sum of these scores for each test item produced the *Darwinist principles score* (range: 0-12).

Note that misconceptions may nevertheless include some explicit reference to Darwinian principles. For example, a Lamarckian account of evolution may refer to part of the *differential survival rates* principle by mentioning the importance of the acquired trait for survival, or to part of the *differential reproduction rates* principle by explicitly stating that the (acquired) trait will be passed on to the next generation,.

One-third of the total of 372 target item responses was coded by two independent human coders, who were blind to the experimental hypotheses. Coding scores were compared and disagreements resolved through discussion. The coders then coded an additional quarter of the total amount of data. The Pearson correlation between raters was r = .97 on the explanatory schema score (r = .96 on the first type and r = .92 on the second type target items) and r = .95 on the Darwinian principles score (r = .95 and r = .93, respectively).

Dialog transcripts. So as to ensure that our manipulation indeed caused a difference in dyadic argumentation we developed a categorization of argumentation in interactions. This coarse categorization is adapted from Michael Baker (1999; 2003). We regarded the dyad as a single unit, the dialogue as a text and categorized each dyadic interaction on the target item (the Duck item) according to one of the following three categories:

(a) *Dialectical argumentation*: The dialogue contains more than one solution which Ss feel obliged to choose from, or the dialogue contains a single proposed solution that is both contested and defended (i.e., Ss consider both the pros and the cons). Note that such interactions do not necessarily have to be adversarial; they may very well be cooperative explorations of different solutions;

(b) *One-sided argumentation*: The dialogue contains only one proposed solution and Ss provide justifications and explanations for it;

(c) *No argumentation*: Ss proposed one or more solutions (detailed or not) without providing any justification or reason for their (in)correctness.

This classification does not consider whether the different solutions or arguments are distributed among the different interlocutors.

To control for differences in the dyadic interaction that could have been caused by the manipulation, other than the type of argumentation, we also investigated the dialogues to see: (1) whether the interaction was of collaborative nature in which both participants contribute equally to the discussion; (2) whether the Darwinist account was mentioned at least once during the interaction; and (3) whether Ss overtly considered the key issue of how the ducks' feet could have changed from "feet like those of pigeons" to "webbed feet".[1]

As a result of technical problems with the audio-recording equipment five dialogues could not be transcribed. Fifty-eight percent of the total number of 33 dialog transcripts were coded by two independent raters. Measures of inter-judge reliability were as follows: *Cohen's Kappa* = .84 for argumentation; *Cohen's Kappa* = .79 for whether the correct solution was mentioned; and *Cohen's Kappa* = .79 for the issue of how the change occurred was discussed. All but two of the interactions were classified as collaborative by both raters; inter-judge reliability could therefore not be calculated.

Results

The assessment yielded the following 10 individual measures: the subject's explanatory schema score at T1, T2 and T3 for each target question (2, 1 and 2, respectively) and the total number of correct Darwinian principles mentioned in response to each of the items. These were then used to calculate two dependent measures: (1) the mean of the explanatory schema scores each subject attained on a given test occasion; and (2) the mean number of Darwinist principles used per item on a given test occasion.

Assessment of item difficulty

A separate sample of 23 undergraduates filled out questionnaires that contained the five target items, in two different orders of presentation. The students were recruited from a course on technology in education and completed the questionnaires as part of their course requirement. Item difficulty was assessed on the two dependent measures: the explanatory schema and the Darwinist principles score in each response (see Coding procedures). The mean explanatory schema scores on the five different test items ranged from $M = 3.02$ ($SD = .93$) to $M = 3.21$ ($SD = 1.16$), whereas the mean number of Darwinist principles subjects mentioned in their responses ranged from $M = 1.60$ ($SD = 2.33$) to $M = 1.91$ ($SD = 2.13$). Item difficulty was not found to differ between test items on either of these measures ($F < 1$). The mean explanatory schema scores were then calculated for the two pretest and the two delayed post-test items. The mean score for pretest items ($M = 3.13$, $SD = 1.09$) was not found to differ from the mean score for post-test items ($M = 3.11$, $SD = 1.61$, $t < 1$). No order effects were found.

Manipulation check and control variables

Chi-square tests for independence were conducted on each of the four dialogue features. The condition and the argumentative nature of the dialogue were related, Fisher's Exact test, $\chi^2 (2, N = 33) = 17.39$, $p < .001$. About two-thirds of experimental dyads engaged in dialectical argumentation and about a third in one-sided argumentation. The majority of the control dyads did not engage in argumentation at all. Interestingly, however, some of the control dyads spontaneously engaged in dialectical argumentation without being prompted to do so and, in spite of the explicit instruction, some of the experimental dyads did not.

The two conditions did not differ for collaboration or for whether the Darwinist answer was mentioned during the interaction (Fisher's Exact test, $\chi^2 (1, N = 33) = 2.77$, *ns* and $\chi^2(1, N = 33) = .25$, *ns*, respectively). However, the condition and the discussion of the key issue of change were related, $\chi^2(1, N = 33) = 4.16$, $p = .04$: A number of experimental dyads greater than chance discussed the crucial issue of how the process of change could have occurred or expressed overt dissatisfaction with that . The opposite was true for the control dyads.

The dyads were also compared for the total number of turns in the dialogue on both test items. The dialogues of the experimental and control dyads were not different on the number of turns, *t* (31) = 1.61, *ns*. Moreover, the number of turns did not correlate with learning gains from T1 to T3 (*r* = -.10, *ns*).

Unit of analysis

Since the design of this study relied on individual assessments and dyadic interventions, we conducted a number of analyses to see whether the axiom of independence was violated. Intra-class correlations were calculated between the two dyad members' explanatory schema scores, on both the pretest and the delayed post-test. There was no correlation between pretest scores, but a significant correlation was found between delayed post-test scores (ρ_I = .44, *p* = .003). A similar pattern was found for the number of Darwinian principles. However, all subjects had seen the instructional movie which presented the Darwinian account of evolution. It is therefore not clear whether this increase in similarity should be attributed to the dyadic interaction, the effect of the movie, or both.

Another method for testing dependency is to see whether the absolute difference between the two partners' mean explanatory schema scores decreased from the pretest to the post-test.[2] If so, that would indicate that the axiom of independence would have been violated. The absolute gap between the pretest scores of the two dyadic partners (*M* = .75, *SD* = 1.04) did not differ from their delayed post-test scores (*M* = .76, *SD* = .81), *t* < 1. However, the lack of a significant change cannot prove that they are independent.

In sum, it is unclear whether the axiom of independence was violated in this particular study. Analyses of individual gains were therefore conducted with adjustments for nested effects (of the individual within the dyad).

The following analyses of learning gains were conducted and are presented in the next section: the effect of condition on the mean improvement from pretest to delayed post-test, the interaction between time and condition, post-hoc analyses of the patterns of change from one test to the next by condition, and an analysis on naturally occurring differences in interaction and conceptual gains.

Conceptual understanding by experimental condition

Four subjects did not complete the delayed post-test. These missing values were replaced by the mean delayed post-test scores of each of the two dependent variables. Analyses were conducted using a mixed model with random effects of dyad within condition and individual within dyad and condition (SAS proc-mixed). Tables 6 and 7 present the adjusted means of the explanatory schema score and the Darwinist principle scores, respectively, for each of the three assessments by condition.

There was an interaction effect of time of assessment and for the explanatory schema score, F (2, 148) = 4.05, p = .019, with an almost negligible effect size, η^2 = .013. A comparison of mean conceptual gains from the pretest to the delayed post-test revealed a significant difference between the experimental and the control condition, F (1, 36) = 4.20, p = .048, with a small effect size of η^2 = .073 (see Table 6).

To examine how students' conceptual understanding changed over time, post-hoc analyses were conducted on improvement from one test to the next for each experimental condition separately. Alpha was adjusted with Tukey-Kramer adjustments for multiple comparisons.

These comparisons showed that, as s a result of the intervention, Ss in both conditions improved their mean explanatory schema score from pretest (T1) to immediate post-test (T2). However, whereas control Ss lost this initial gain and dropped back to their initial level of performance at the delayed post-test a week later (T3), the mean explanatory schema score of experimental Ss only slightly decreased, a change that did not reach significance (see Table 6). Effect sizes ranged from medium to large.

For the mean Darwinist principles scores, on the other hand, there was no interaction effect, $F < 1$, and the two conditions did not differ in their respective mean difference scores from T1 to T3, $F < 1$ (see Table 7). Post-hoc analyses revealed that immediately following the intervention, Ss in both conditions increased the number of Darwinian principles in their responses. However, this increase was not maintained at the post-test occasion a week later, as subjects in both conditions decreased the number of Darwinist principles in their T3 responses, even though this decrease seems smaller in the experimental than in the control condition.

Conceptual gains and correlative differences in argumentation

Table 8 presents the dyads' mean explanatory schema model score and mean Darwinist principles score on each of the tests by type of dyadic interaction (see Method section): no argumentation, one-sided, or dialectical argumentation. Dyads that engaged in dialectical argumentation showed gains from pretest to delayed post-test, while dyads whose interaction was non-argumentative or one-sided did not improve from T1 to T3. The patterns of change were quite similar to the ones found in the explanatory schema score analysis by experimental condition: Dyads who engaged in dialectical argumentation gained slightly more at T2 and were able to retain this gain, whereas the other dyads gained less at T2 and regressed at T3. Due to the relatively small and unequal number of dyads in cells, we conducted non-parametric Kruskal-Wallis tests on improvement from pretest to delayed post-test by type of dyadic interaction. On the explanatory schema score, improvement from T1 to T3 was found to be dependent on the nature of interaction, χ^2 $(2, N = 33) = 10.71$, $p = .005$. No dependency was found on the Darwinist principle score, χ^2 $(2, N = 33) = 3.01$, *ns*. Thus, the analyses of naturally occurring differences in argumentation seem to mirror the experimental effects.

Discussion

In this study we found that students who were instructed to engage in dialogical argumentation with a peer showed greater gains in understanding evolutionary concepts than control students. Moreover, changes in conceptual understanding were found to be dependent on experimental condition: Whereas students in both conditions improved on the explanatory schemas immediately following the intervention, only students in the experimental condition preserved these gains until the delayed post-test a week later. Moreover, analyses on naturally occurring differences in type of interaction seemed to indicate that dialectical argumentation was crucial for gains in conceptual understanding.

The improvement in conceptual understanding as seen in the explanatory schemas students used could not be attributed to an increase in the number of correct Darwinian principles they produced: Students in both conditions showed immediate gains which disappeared on the delayed post-test a week later.

There is some research that seems to corroborate the Piagetian claim that the cognitive restructuring that follows disequilibration can take some time to materialize since intra-mental processes and knowledge need to be recalibrated (Azmitia & Crowley, 2001). For example, in their research on peer collaboration and conceptual change in physics, Howe and her associates (Howe, McWilliam & Cross, 2005; Howe, et al., 1992; Tolmie, Howe, McKenzie & Greer, 1993) showed that slight-to-moderate cognitive gains at the immediate post-test increased significantly at the delayed post-test.

Similar to Howe's findings, mentioning the correct solution during the interactions in our study was not found to be dependent on experimental condition. However, our data did not show a pattern of delayed gains, but rather *preserved* gains for those who were instructed to engage in dialectical argumentation, and loss of temporary gains for those who were not. Moreover, the long-term advantage of the argumentative condition was on the explanatory schema level only.

These findings, combined with the fact that the two groups showed similar levels of performance on the immediate post-test, suggest that the differences in conceptual understanding may be the result of different levels of processing during or after the intervention phase: The engagement in argumentation may have led to superior processing which resulted in consolidation of their immediate cognitive gains. On the post-test occasion, these restructured cognitive schemas were then retrieved and applied to new phenomena. In contrast, the immediate gains of subjects that did not engage in argumentation may have been due to the social affordances of the collaborative situation (such as exposure to superior solutions from more knowledgeable or dominant partners) or recall of information from the movie.

Our conjecture that the difference between the two conditions may be attributed to superior processing as a result of argumentation, would be further supported if such patterns of change could be replicated in an additional study. The next experiment was designed to address these and other issues in a more rigorously controlled design. The small effect sizes for the Study 1 effects of condition on conceptual gains imply that much of the variance is not accounted for. It is likely that this is in part due to the peer collaboration nature of the experiment, which caused undesired intra-group differences in dyadic argumentation. Therefore, some changes were made in the interventions to further isolate the engagement in dialectical argumentation. To examine whether argumentation leads to superior processing *during* or *after* the intervention, students were also asked to report post-intervention reflections on evolution.

In addition, Study 2 also tested the possibility that the effects in the first study may have been caused by the critical discussion excerpt that was presented only to the experimental subjects. Even though the item was aimed at modeling a critical discussion and the correct solution was not hinted at, the excerpt contained some content information, and we cannot refute the possibility that it might have been responsible for some of the conceptual gains during and following the interaction. Therefore, the excerpt was omitted in the next study.

Study 2: Monological Argumentation

A confederate played the role of one of the participants in both conditions. Participants in the experimental condition were prompted to engage in dialectical argumentation on their own and the confederate's solution, by answering structured questions read aloud by the confederate, who chose a piece of paper from an urn and invariantly picked up the role of the "reader". In the control condition, the subject and the confederate only read aloud their solutions to each other, without discussing them further. Thus, students in both conditions were prevented from conducting a natural dialogue and were exposed to the same naïve conception in evolution.

The task scenario in the experimental condition was designed to ensure that participants engaged in dialectical argumentation while preserving the perceived equal-status, peer-collaborative nature of the task in Study 1. The choice of using a confederate instead of assigning two participants to different roles was motivated by two reasons: First of all, the confederate could make sure that Ss would not skip any of the questions that induced them to engage in dialectical argumentation. Secondly, the new experimental setting avoids any questions of the appropriate unit of analysis of individual cognitive gains.

In both conditions, Ss interacted with the confederate and the additional solution was always presented as being the confederate's, who personally read it to Ss. In sum, the conditions were identical on factors such as social facilitation, source of content (whether provided by the dyadic partner or the dialog excerpt in the first study), exposure to alternative views and personification of viewpoints, and differed only in engagement of argumentation.

Method

Participants

Forty-four university students from the Education (10%), Social Sciences (56%) and Humanities (34%) departments were recruited in the same manner as in Study 1 (mean age = 27.14). The data of two Ss were omitted from the analyses since they did not finish the pre-test questionnaires within the 25 minutes time limit. Most Ss (76%) reported not having received any formal education on the subject of evolution in either high school or academic courses. Forty participants were financially rewarded, two received course credit.

Materials

Pre- and post-test materials were similar to Study 1. After the evolutionary items on the post-test, Ss were asked whether they had thought about the topic of evolution or any aspect of the experiment in the week between the intervention and the post-test. They were also asked whether they had consulted with any human or non-human source on the subject. The two-page booklet used in the experimental intervention contained general instructions concerning the task and nine steps to be read aloud by the confederate (see Procedure for further details). The warm-up item of the Study 1 intervention was omitted.

Procedure

Except for the interventional phase, the procedure and materials were identical to Study 1. Immediately after watching the movie excerpt, Ss were asked to answer the intervention target item used in Study 1 (the Duck item) and hand in their answer sheets. They were then introduced to the confederate who, they were told, had just completed the same stages of the experiment as they had. The confederate was a female undergraduate of the same age group as the participants. Great efforts were made to ensure that her physical appearance and behavior did not reveal that she was confederate.

Ss in the experimental conditions were told that they would participate in a short collaborative task with clearly defined roles and instructions. The confederate was then "randomly" assigned the task of reading aloud the instructions and questions from a booklet, whereas the participant was always "randomly" assigned to answer these questions. All interactions were audio-taped.

The confederate then read aloud the general instructions which briefly described the task and stated that the goal of the activity was to gain a better understanding of evolutionary change processes, rather than to win a certain argument. The task itself consisted of nine steps that were read aloud by the collaborator. Six of these steps prompted the participant to engage in dialectical argumentation on their own and the confederate's solution to the Duck item; three of these steps contained procedural instructions.

First, participants were requested to read aloud their answer to the duck item from the previously filled-out answer sheet. They were then asked to discuss the strengths of that solution, to criticize it, and to discuss whether it explained the change that occurred to the ducks' feet. Then the confederate was requested to read "her" solution aloud, after which the participants were asked to discuss that solution according to the previous steps. Finally, they were requested to evaluate the two solutions and their personal conceptual understanding of evolution. At each of the six argumentative steps the collaborator read the question aloud and waited for participants to finish their response, after which she then prompted them to elaborate some more ("Why do you think so? Is there anything else you would like to add?"). At the end of the collaborative task, confederate and participant separately answered the duck item one more time on a fresh answer sheet (T2). Table 9 presents the full description of the prompts read by the confederate.

The confederate was instructed to strictly adhere to reading the written prompts and to react neutrally, but empathically to the participants' answers. She was instructed not to be responsive to participants' efforts to elicit a dialogue. In such cases, she would appear hesitant but then draw their attention to the task instructions according to which she could only read the prompts.

So as to ensure uniformity of exposure to another explanatory schema (i.e., the confederate's), while preserving a minimum difference between that and the S's explanatory schema, two different answer sheets were prepared for the confederate: One containing a typical Lamarckian answer and the other a typical answer from the "mutations after" schema (see Table 3). Both models belong to the typological change category, in which intra-species variation is not considered and all individuals within a group undergo the transformation (typological thinking). They differed, however, in their description of the way the change occurred: Acquired traits are inherited by offspring (Lamarckian) or individuals undergo "mutations" in their lifetime (mutations after).

When the dyad started reading the instructions, the subject's explanatory schema was quickly assessed based on his/her answer immediately following the movie. The confederate's response was then chosen according to the following principle: If the subject's response represented a Lamarckian schema, the confederate was given the "mutations after" response. In any other case, she was given the Lamarckian response. The experimenter then re-entered the room and "returned" the dyad's answer sheets in time for the subject to read aloud his/her answer, with the excuse that the experimenter had forgotten to return them. The average length of the scripted dialog was 8:05 minutes (ranging from 5:50 to 13:40).

In the control condition, Ss were also seated with the confederate. They were first given a short neutral task (eight questions on which animal is the fastest, strongest, heaviest, and so on, alive today) for which they had seven minutes to complete. During this task the experimenter assessed the participant's explanatory schema. Their sheets were then "returned" to them and they were requested to read their respective answers to each other while refraining from any further verbal communication.

The confederate was given the Lamarckian response in 46% of the experimental sessions and in 50% of the control cases. Immediately following the delayed post-test (T3), each student was briefed on the confederate's identity. All Ss were reported to be surprised and stated they did not suspect anything of the kind during the experiment.

Coding procedures

Coding procedures for assessing evolutionary understanding were identical to those presented in Study 1. To test the manipulation's effectiveness, protocols of the conversations in the experimental condition were transcribed and coded according to the following criteria: number of reasons participants proposed in favor of their own solution (strengths of their solution), number of objections they proposed against their solution (weaknesses in their solution), and number of supporting and objecting arguments for and against the confederate's solution. Repetitions and supports or objections that did not discuss content, but were purely evaluative (e.g., "Your solution is better than mine") or that discussed only superficial features of the solution (e.g., "It is formulated very well") were discarded. Any support or objection that did involve content was coded, whether it was objectively correct or not. Two independent human coders coded all twenty-two transcripts. Pearson correlations between their assessments were r = .81 (support own), r = .97 (objections own), r = .78 (support other) and r = .84 (objections other). Disagreements were resolved through discussion.

Results

Manipulation check

Nineteen of the 22 experimental subjects proposed at least one supporting reason for their own solution and an identical number of students proposed at least one objection. The number of subjects who mentioned at least one argument for or against the confederate's solution was smaller (12 and 16, respectively). Two subjects provided support only for their own solution without further criticizing their own or the confederate's solution. All other subjects engaged in dialectical argumentation. The mean number of supporting arguments for one's own argument (*M* = 1.22, *SD* = .75), objections to own (*M* = 1.27, *SD* = .83), supporting arguments for other (*M* = .59, *SD* = .59) and objections against other (*M* = .91, *SD* = .87) were all significantly greater than zero ($p < .001$).

Conceptual gains

About half of the participants in each condition reported that they had thought about evolution in the week following the intervention ($\chi^2 < 1$) and only three individuals reported that they had consulted with friends without receiving answers from them.

The analyses were similar to those conducted in Study 1. The mean explanatory schema score and the mean number of Darwinist principles were analyzed in two separate analyses of variance (2 X 3) with time of assessment (pretest vs. immediate post-test vs. delayed post-test) as a within-subject factor, condition as a between-subject factor and formal education in biology as a covariate. Huyn-Feldt corrections were applied when necessary. Previous education in biology did not have any effect in either of the analyses ($F < 1$ and F (2, 78) = 1.96, *ns*, respectively). Figures 1 and 2 plot the unweighted adjusted means for the explanatory schema score and the Darwinist principles score, respectively, on the three test occasions by condition.

There was an interaction effect for test occasion and condition on the explanatory schema score, F (1.70, 66.42) = 3.70, $p = .037$, with a small-to-medium effect size of η^2 = .09. Whereas experimental subjects improved their conceptual understanding from the pretest to the delayed post-test by more than half a point (*M* = .55, *SE* = .18), control subjects did not improve at all (*M* = -.08, *SE* = .17), F (1, 39) = 6.07, $p = .018$, with a medium effect size of η^2 = .14.

To examine whether the learning patterns found in the first study could be replicated, planned comparisons were conducted between the three different test occasions (from T1 to

T2 and from T2 to T3) within each condition (see Figure 1). As in Study 1, Ss' cognitive gains in the experimental condition were obtained immediately following the intervention (F (1, 20) = 4.60, p = .044, η^2 = .19) and a similar level of performance was maintained on the delayed post-test (F < 1). However, unlike the results of the control condition in Study 1, control subjects in this study showed a small and non-significant improvement at the immediate post-test (F (1, 18) = 1.89, *ns*), followed by a small and non-significant decrease at the delayed post-test (F (1, 18) = 2.00, *ns*). In other words, the explanatory schema score of control subjects did not significantly change on the different test occasions.

Unlike Study 1, an interaction between condition and test occasion was also found for the mean Darwinist principles score, F (1.70, 66.32) = 3.63, p = .039, with a small-to-medium effect size of η^2 = .09. Comparisons between learning gains from pretest to delayed post-test revealed an advantage of experimental (M = 1.01, SE = .41) over control subjects (M = -.64, SE = .43), F (1, 39) = 7.37, p = .010, with a medium effect size of η^2 = .16. Figure 2 shows the following patterns of change: In the experimental condition, the pattern of change over the three test occasions resembles the trends of immediate and maintained improvement for the explanatory schema score (see Figure 2), whereas control subjects seem to have lower Darwinist principles scores on each succeeding test. However, none of the four planned comparisons reached significance.

Discussion

The results presented here provide further support for the hypothesis that argumentation promotes conceptual understanding in an instructional design based on the cognitive conflict paradigm. Students who were instructed to engage in dialectical monological argumentation on their own and another person's solution showed greater conceptual gains than control students. Thus, the advantage of argumentation observed in collaborative dyadic situations was replicated in a situation of monological argumentation directed at and prompted by a peer, a design that isolated the argumentation from the collaborative features of the interaction. Moreover, students in the experimental condition also showed larger gains on the use of Darwinist principles than control subjects.

As in Study 1, students in the argumentative condition preserved the conceptual gains obtained during the intervention, while students in the control condition did not show improvement on any of the tests. This suggests that control subjects' temporary gains in Study 1 derived from the collaborative nature of the task, and not from the movie they saw. When students were not allowed to interact and were given the same misconception, such temporary gains disappeared.

Since instances of post-intervention reflection and deliberation were equally distributed among experimental and control students, the effects could not be attributed to these variables.

General Discussion

Argumentation has become an increasingly popular subject of investigation in the last 15 years or so. Testimony to this upsurge in interest may be found in a number of fields of expertise: argumentation theory itself, where models have been developed to distinguish argumentation from other forms of reasoning (e.g., Van Eemeren, Grootendorst, Henkenmans, Blair, Johnson, Krabb, Plantin, Walton, Willard, Woods & Zarefsky, 1996); discourse analysis, where coding systems were elaborated for the assessment of argumentative discourse (e.g., Kuhn, 1991; Leitao, 2000; Resnick, Salmon, Zeitz, Wathen & Holowchak, 1993); psychology, where models have been developed to stress the importance of argumentation in (critical) thinking processes and the organization of knowledge representations (e.g., Billig, 1996; Moshman, 1998); and developmental psychology, to observe argumentative skills in different age groups (e.g., Felton & Kuhn, 2001; Glassner & Schwarz, 2005; Kuhn, 1991; Voss & Means, 1991). However, although argumentation is often associated with knowledge construction, empirical research has seldom focused on the benefits of argumentation for learning (Duschl & Osborne, 2002). The two studies conducted here will hopefully be a step in this direction, for concept learning.

This study addressed the question of whether engagement in dialectical argumentation promotes concept learning, particularly in learning situations that implement cognitive conflict techniques. The hypothesis was tested in two different formats of argumentation. In the first study, dyads were instructed to conduct a critical argumentative discussion while solving evolutionary problems together. Students in the second study engaged in monological dialectical argumentation on their own and another person's solution in reaction to structural argumentative prompts read aloud by a confederate. In both formats, students who participated in argumentation-eliciting conditions showed greater gains in understanding of evolutionary concepts than control subjects. Moreover, these learning gains were dependent on naturally occurring differences in dialogical argumentation: Only dyads that engaged in dialectical argumentation showed gains at the delayed post-test; dyads that did not engage in argumentation or that engaged only in one-sided argumentation did not.

The findings reported here suggest that engaging in dialectical argumentation indeed promotes concept learning in collaborative learning situations, at least for evolution. Studies by King (1990; King & Rosenshine, 1993) and Coleman (1998) have shown that altering the discourse in collaborative learning settings can optimize their effectiveness. The type of discourse they investigated, however, emphasized explanation-driven interactions. The findings presented in this study seem to extend theirs, by identifying argumentation as a condition for effective collaborative learning.

The effects of argumentation found in this study are particularly noteworthy given the short intervention students underwent and given the fact that the concept of evolution is known to be particularly resistant to change (e.g., Brumby, 1984; Ohlsson, 1992; Ohlsson &

Bee, 1992). Instructional interventions in authentic educational settings are likely to be more intensive and have a larger impact. Dialectical argumentation requires explaining oneself and justifying one's standpoints, as well as dialectically discussing and evaluating different solutions. We proposed that the advantage of argumentation for conceptual understanding is achieved through the superior processing of information. This conjecture is indirectly supported by two findings: (1) mentioning the Darwinian account during the discussion was not related to learning gains; and (2) the particular pattern in which the advantage of argumentative conditions was achieved in both experiments (preserved gains versus loss of temporary gains or no gains). These findings also emphasize the importance of assessing learning gains at different time intervals.

Our findings concerning conceptual understanding in evolutionary theory are intriguing. One interesting finding is that the patterns of improvement in explanatory schemas were not always paralleled by the use of correct Darwinian principles. Our findings seem to echo those presented in Chi and Roscoe (2002), in that reorganizations of knowledge (i.e., changes at the explanatory schema level) are not necessarily accompanied by equal increases in the number of correct propositions individuals produce. Wiley and Voss (1999) found that argumentative essay writing led to an increase in conceptual understanding but not in factual knowledge. Based on our findings and somewhat similar to those reported by Wiley and Voss, argumentation seems to promote changes at the explanatory schema level, but not always on the propositional level. However, this conclusion may prove to be limited to the open question format used in this study.

It is well known that argumentation on scientific issues does not occur naturally (de Vries, Lund & Baker, 2002). Indeed, the large majority of undergraduate dyads did not engage in this type of dialogue when they were not specifically instructed to do so. In other words, designing learning tasks according to the traditional cognitive conflict paradigm does not suffice, even in the select sample of university undergraduates. Students have to be specifically instructed and encouraged to engage in dialectical argumentation on a scientific notion. This finding adds to other studies that have emphasized the importance of structure in cooperative learning activities (Gillies, 2003; King & Rosenshine, 1993).

Our findings provide first experimental evidence of conceptual learning gains in the sciences as a result of verbal argumentation. More work is needed to establish the criteria for the implementation of argumentation in authentic learning contexts. In addition, further research is needed to generalize our findings to other learning settings, populations and subject matter and to explore the mechanisms through which argumentation improves conceptual understanding.

References

Amigues, R. (1988). Peer interaction in solving physics problems: Socio-cognitive confrontation and meta-cognitive aspects. *Journal of Experimental Child Psychology, 45,* 141-158.

Azmitia, M. A. & Crowley, K. (2001). The rhythms of scientific thinking: A study of collaboration in an earthquake microworld. In K. Crowley, C. Schunn, & T. Okada (Eds.) *Designing for science: Implications from everyday, classroom, and professional settings.* Mahwah, NJ: Erlbaum.

Baker, M. (2003). Computer-mediated interactions for the co-elaboration of scientific notions. In: J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments*. Utrecht: Kluwer Academic Publishers.

Baker, M. (1999). Argumentative interactions, discursive operations, and learning to model in science. In: P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and computational approaches.* Amsterdam: Pergamon.

Billig, M. (1996). *Arguing and thinking. A rhetorical approach to social psychology (2nd Ed).* Cambridge: Cambridge University Press.

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

Brumby, M. N. (1984). Misconceptions about the concept of natural selection by medical biology students. *Science Education*, *68*, 493-503.

Chi, M. T. H. (2000). Self-explaining expository texts: The dual process of generating inferences and repairing mental models. In: Glaser, R. (Ed), *Advances in Instructional Psychology*. Mahwah, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H., Bassok, M., Lewis, M.W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.

Chi, M. T. H., deLeeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439-477.

Chi, M. T. H., & Roscoe, R. D. (2002). The processes and challenges of conceptual change. In: M. Limon & L. Mason (Eds), *Reconsidering conceptual change: Issues in theory and practice.* Utrecht: Kluwer Academic Press.

Coleman, E. B. (1998). Using explanatory knowledge during problem solving in science. *Journal of the Learning Sciences, 7* , 387-427.

De Vries, E.,, Lund, K., & Baker, M. (2002). Computer-mediated epistemic dialogue: Explanation and argumentation as vehicles for understanding scientific notions. *Journal of the Learning Sciences, 11*, 63-103.

Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education, 38,* 39-72.

Felton, M., & Kuhn, D. (2001). The development of argumentative discourse skill. *Discourse Processes, 32*, 135-153.

Ferrari, M. & Chi, M. T. H. (1998). The nature of naive explanations of natural selection. *International Journal of Science Education, 20*, 1231-1256.

Gillies, R. M. (2003). The behaviors, interactions and perceptions of junior high school students during small-group learning. *Journal of Educational Psychology, 95*, 137-147.

Glassner, A. & Schwarz, B. B. (2005). The Antilogos Ability to evaluate information supporting arguments. *Learning and Instruction, 15*, 353-375.

Greene, E. D. (1990). The logic of university students' misunderstanding of natural selection. *Journal of Research in Science Teaching, 27,* 875-885.

Howe, C., McWilliam, D., & Cross, G. (2005). Chance favors only the prepared mind: Incubation and the delayed effects of peer collaboration. *British Journal of Psychology*, *96*, 67-93.

Howe, C., Tolmie, A., Duchak-Tanner, V., & Rattay, C. (2000). Hypothesis-testing in science: Group consensus and the acquisition of conceptual and procedural knowledge. *Learning & Instruction*, *10,* 361-391.

Jensen, M. S., & Finley, F. N. (1996). Changes in students' understanding of evolution resulting from different curricular and instructional strategies. *Journal of Research in Science Teaching, 33*, 879-900.

Jimenez-Aleixandre, M. P. (1992). Thinking about theories or thinking with theories?: A classroom study with natural selection. *International Journal of Science Education, 14,* 51-61.

King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal questioning. *American Educational Research Journal, 27*, 664-687.

King, A. & Rosenshine, B. (1993). Effects of guided cooperative questioning on children's knowledge construction. *Journal of Experimental Education, 61,* 127-148.

Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.

Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review, 62*, 155-178.

Kuhn, D., & Lao, J. (1998). Contemplation and conceptual change: Integrating perspectives from social and cognitive psychology. *Developmental Review, 18,* 125-154.

Kuhn, D., Shaw, V. & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. *Cognition & Instruction, 15*, 287-315.

Leitao, S. (2000). The potential of argument in knowledge building. *Human Development, 43*, 332-360.

Limon, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning & Instruction, 11*, 357-380.

Mason, L. (2001). Introducing talk and writing for conceptual change: a classroom study. *Learning & Instruction, 11*, 305-329.

Means, M. L. & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition & Instruction, 14*, 139-179.

Moshman, D. (1998). Cognitive development after childhood. In: W. Damon (Series Ed) and D. Kuhn (Vol. Ed), *Handbook of child psychology, vol. 4 (5th Ed),* (947-978). New York: Wiley.

Neuman, Y. & Schwarz, B. B. (2000). Substituting one mystery for another: The role of self-explanations in solving algebra word-problems. *Learning & Instruction, 10*, 203-220.

Nussbaum, E. M., & Sinatra, G. M. (2003). Argument and conceptual engagement. *Contemporary Educational Psychology*, *28*, 384-395.

Ohlsson, S. (1992). *Young adults' understanding of evolutionary explanations: Preliminary observations*. Technical Report. Learning Research and Development Center, University of Pittsburgh.

Ohlsson, S. (2002). Generating and understanding qualitative explanations. In: J. Otero, J.A. Leon, & A.C. Graesser (Eds), *The psychology of science text comprehension* (91-128). Mahwah, NJ: Lawrence Erlbaum.

Ohlsson, S. & Bee, N. V. (1992). *The effect of expository text on students' explanations of biological evolution.* OERI Report. Learning Research and Development Center, University of Pittsburgh.

Okada, T. & Simon, H.A. (1997). Collaborative discovery in a scientific domain. *Cognitive Science, 21*, 109-146.

Resnick, L. B., Salmon, M., Zeitz, C. M., Wathen, S. H., & Holowchak, M. (1993). Reasoning in conversation. *Cognition & Instruction, 11*, 347-364.

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

Sandoval, W.A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Sciences, 12,* 5-51.

Schnotz, W., & Preuss, A. (1999). Task-dependent construction of mental models as a basis for conceptual change. In: W. Schnotz, S. Vosniadou & M. Carretero (Eds), *New perspectives on conceptual change* (193-222). Amsterdam: Pergamon Press.

Schwarz, B. B. (2003). Collective reading of multiple texts in argumentative activities. *The International Journal of Educational Research, 39,* 133-151.

Schwarz, B. B. & Glassner, A. (2003). The blind and the paralytic: Supporting argumentation in everyday and scientific issues. In: J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments.* Utrecht: Kluwer Academic Publishers.

Schwarz, B. B., Neuman, Y., Gil, J., & Ilya, M. (2003). Construction of collective and individual knowledge in argumentative activity. *The Journal of the Learning Sciences, 12*, 221-258.

Schwarz, B. B., Neuman, Y., & Biezuner, S. (2000). Two wrongs may make a right...if they argue together! *Cognition & Instruction, 18,* 461-494.

Tetlock, P. (1992). The impact of accountability on judgment and choice: Toward a social contingency model. In: M. Zanna (Ed.), *Advances in experimental social psychology, Vol. 25.* San Diego, CA: Academic Press.

Tolmie, A., Howe, C., McKenzie, M, & Greer, K. (1993). Task design as an influence on dialogue and learning: Primary school group work with object flotation. *Social Development, 2,* 183-201.

Van Eemeren, F. H., Grootendorst, R., Henkenmans, F. S., Blair, J. A., Johnson, R. H, Krabb, E. C., Plantin, C., Walton, D. N., Willard, C. A., Woods, J. & Zarefsky, D. (1996). *Fundamentals of Argumentation theory: A handbook of historical background and contemporary developments.* Hillsdale, NJ: Erlbaum.

Vosniadou, S. & Brewer, W. F. (1994). Mental models of the day / night cycle. *Cognitive Science, 18,* 123-183.

Voss, J. F. & Means, M. L. (1991). Learning to reason via instruction in argumentation. *Learning & Instruction, 1,* 337-350.

Wiley, J. & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91,* 301-311.

Author Note

Christa S. C. Asterhan and Baruch B. Schwarz, School of Education, Hebrew University of Jerusalem.

This research was supported in part from the Israeli Foundations Trustees, Grant 27.35 and is based on a doctoral dissertation by Ms. Asterhan.

Correspondence concerning this article should be addressed to Christa Asterhan, School of Education, Hebrew University of Jerusalem 91905, Israel. Email: kooij@mscc.huji.ac.il

Footnotes

[1] Giving a Darwinian account is not required on this dimension, since it merely assesses whether they gave the issue some thought or not.

[2] In all but two dyads the direction of the difference between two partners' mean explanatory schema scores remained constant.

[3] We do not consider the differences between Lamarckian, teleological and other similar types of explanations here, since they all share the notion that individuals are considered to acquire traits and to pass these on to their offspring.

Table 1

Examples of the Three Different Types of Test Items

Test item	Example
Warm up	In the tropical rain forests of Panama, Central America there is a species of butterfly called the Amartia Fatima. The butterflies have different color patterns: some have striped wings whereas others have only one color. Recently, a change in the extent of wing damage of these butterflies has been found: The wings of the striped ones have been found to be more damaged than those of the plain-winged butterflies of this species. It is also known that this damage is caused by the attacks of birds. Given these facts, what change can be expected to occur to the population of Amartia Fatima butterflies in this area?
Target: type 1	DDT is a chemical substance used to kill a number of insects, among them mosquitoes. Some twenty years ago, when the insecticide was first introduced, it was found to be extremely effective in killing mosquitoes. However, today a number of mosquito populations have been found to be resistant to DDT and a new insecticide will have to be developed for them. Given these facts, how do you think evolutionary theory would account for the process of change that caused for so many mosquitoes to be resistant today?
Target: type 2	When chasing prey, a cheetah can run as fast as 95 km/hour, which is faster than any other predator in his habitat. However, the cheetah's ancestors, who lived thousands of years ago, could reach a maximum speed of "only" 32 km/hour. Given these facts, how do you think evolutionary theory would account for the process of change that occurred in the running speed of the cheetah?

Table 2

The Evolutionary Phenomena tested on each of the Three Test Occasions as a Function of Test Item Type

Type of test item	Pretest	Immediate post-test	Delayed post-test
Warm-up item	Butterflies (wing pattern)	Finches (beak size)	Butterflies (wing pattern)
Target item: type 1	Mosquitoes (resistance)	Ducks (webbed feet)	Moths (coloration)
Target item: type 2	Cheetahs (running speed)		Sea Iguanas (swimming ability)

Table 3

Characteristics of the Six Explanatory Schema Categories for Evolutionary Change

Schema category	Description
Non-answers	Responses that indicated absolute ignorance on the subject (“I dunno”) or that merely repeated the data that was given in the question item.
No change considered	Responses that simply denied that species evolve over time, such as disaster scenarios (“everything will go extinct") or that refer to species intentionally moving away to other (better) areas to protect themselves.
Unexplained change	Responses that refer to species changing over time without providing a reason for it or describing how it occurs. Typically, they refer to evolution as a process according to which everything changes for the better and species continue perfecting themselves.
Typological change	Each of the four schemas in this category has a different account of *why* or *how* this change occurs: Lamarckian (individual members acquire a trait which is then passed on to offspring);[3] Mutations after (as a result of the change in the environment, individual members “undergo mutations” which are then passed on to offspring); Mating outside species (individual members will try to mate with members of other species that have an advantageous trait); and Dormant genes (the trait was always present in the species’ genetic makeup, but was dormant until it was “needed” and activated).
Hybrid explanations	Responses that integrate features of natural selection with typological change mechanisms. They are similar to the former category in that they do not consider existing intra-species variation, but instead refer to variation that is created in reaction to environmental change and needs. However, they do consider selection: The transformed individual members of the species manage to survive and reproduce, whereas the others go extinct. These schemas are often surprisingly coherent and occur fairly often in our sample. Two types of hybrid models occurred, with different accounts of how variation is created: (a) Some individual members of the species underwent genetic mutations in reaction to the change in the environment; and (b) Some individual members acquired the trait and this acquired trait was passed on.
Darwinist	Responses that explain change in terms of natural selection and existing intra-species variation.

Table 4

Prevalence of the Four Defining Dimensions per Explanatory Schema Category

Score	Type of solution	Change of species	Explanation of change	Selection mechanism	Intra-species variation
0	Non-answers	no	no	no	no
1	No change considered	no	no	no	no
2	Unexplained change	yes	no	no	no
3	Typological change	yes	yes	no	no
4	Hybrid explanation	yes	yes	yes	no
5	Darwinian	yes	yes	yes	yes

Table 5

Six Darwinist Principles

Darwinist principle	Description
Intra-species variability	Individuals in one generation of a particular species differ from each other along a number of dimensions.
Source of variability	This variability is the result of random changes in genetic material (sexual recombination and random mutations).
Differential survival rates	Different characteristics are more or less advantageous in a given environment and resources in each environment are limited. Those with the more advantageous characteristics will have an increased chance for survival in that environment.
Reproductive advantage	Individuals with advantageous traits reproduce more and are therefore more likely to pass on their genes to the next generation.
Accumulation of change	When this process is repeated over many generations, the accumulated changes can lead to changes in the population.
Changes within the population	As a result, the proportion of individuals carrying the advantageous trait(s) will increase within the population, whereas the proportion of those without the trait(s) will decrease.

Table 6

Adjusted Means of Students' Explanatory Schema Scores by Test Occasion and Condition

Condition	*M*	*SE*	*t(148)*	*p*	*Cohen's d*
Experimental (*n* = 38)					
T1: Pretest	3.29	.17			
			3.71	.004	.60
T2: Immediate post-test	3.79	.17			
			-.90	ns	-
T3: Delayed post-test	3.66	.17			
Control (*n* = 38)					
T1: Pretest	3.36	.17			
			3.91	.002	.63
T2: Immediate post-test	3.87	.17			
			-4.52	.001	.73
T3: Delayed post-test	3.28	.17			

Table 7

Adjusted Means of Students' Darwinist Principles Score by Test Occasion and Condition

Condition	*M*	*SE*	*t(148)*	*p*	*Cohen's d*
Experimental (*n* = 38)					
T1: Pretest	2.40	.51			
			5.65	<.001	.92
T2: Immediate post-test	4.34	.51			
			-2.90	.048	.47
T3: Delayed post-test	3.35	.51			
Control (*n* = 38)					
T1: Pretest	2.66	.51			
			4.89	<.001	.79
T2: Immediate post-test	4.34	.51			
			-3.73	.004	.61
T3: Delayed post-test	3.06	.52			

Table 8

Mean Explanatory Schema and Darwinist Principles Scores of Dyads by Test Occasion and Type of Argumentation

	Explanatory schema score		Darwinist principles score	
Type of argumentation	*M*	*SD*	*M*	*SD*
None (n = 11)				
T1	3.56	.52	3.25	1.84
T2	3.91	.83	4.32	3.16
T3	3.36	1.03	3.55	2.15
d_{13}[a]	-.19	.90	.30	1.82
One-sided (n = 8)				
T1	2.97	.31	1.16	.56
T2	3.19	.37	2.63	1.21
T3	2.80	.57	1.66	1.33
d_{13}[a]	-.17	.41	.50	1.06
Dialectical (n = 14)				
T1	3.32	.59	2.54	1.71
T2	4.15	.61	5.18	2.08
T3	3.88	.76	3.86	1.87
d_{13}[a]	.55	.50	1.20	1.12

[a] d_{13} is the difference between the mean score at T1 and the mean score at T3.

Table 9

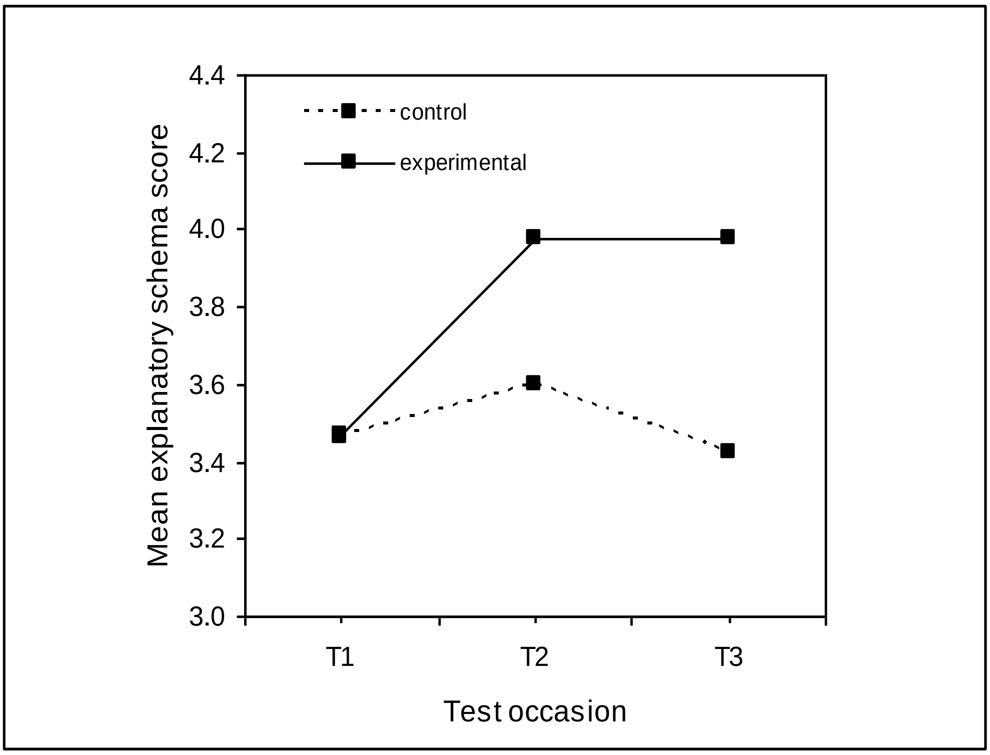
Prompts of the Study 2 Experimental Condition.

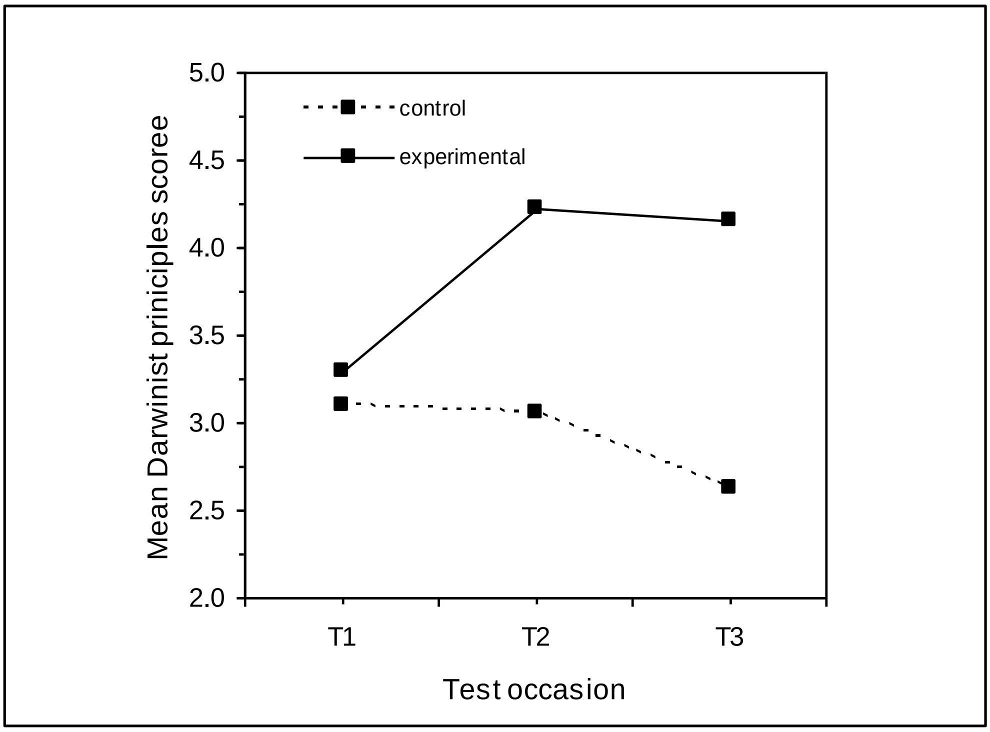
Number	Content
1	Please read me the answer you wrote down on the duck question.
2	I am not sure whether that really explains the duck phenomena, or maybe I did not understand you well. Why do you think that this is the correct answer? In other words, what are your justifications, what is the strength of the solution you proposed?
3	Okay. Now try to take a critical stance towards the solution you proposed. What could be the weak points in this explanation? What elements could be considered unresolved or illogical?
4	In the duck item we were asked to explain the mechanism of change that occurred in the ducks' feet. Does your solution satisfy this requirement? How does it explain the process of change?
5	Now I will read you the solution I wrote.
6	What do you think about this solution? (repeat 2-4 if necessary)
7	.After all we just discussed, what is your opinion on the matter now? I mean, how did the change in the ducks occur?
8	Do you have any doubts concerning the correct explanation, or evolution in general? Are there still any elements that remain unresolved or unclear to you?
9	Now we will both answer the duck question one more time on the fresh answer sheets each of us received. Please answer in a detailed, explicit and elaborated manner, discussing all the relevant ideas and concepts.

Figure Captions

Figure 1. Mean explanatory schema score by test occasion and condition.

Figure 2. Mean Darwinist principles score by test occasion and condition.


4.4
4.2
4.0
3.8
3.6
3.4
3.2
3.0
control
experimental
Mean explanatory schema score
T1
T2
T3
Test occasion


5.0
4.5
4.0
3.5
3.0
2.5
2.0
control
experimental
Mean Darwinist priniciples scoree
T1
T2
T3
Test occasion