PERGAMON

New Ideas in Psychology 20 (2002) 263–283

NEW IDEAS IN PSYCHOLOGY

www.elsevier.com/locate/newideapsych

Science undergraduates' and graduates' epistemologies of science: the notion of interpretive frameworks ☆

Claudia Thoermer*, Beate Sodian

Department of Developmental Psychology, Ludwig-Maximilians-University, Leopoldstr 13, D-80802 Munich, Germany

Abstract

First-year science (biology, chemistry, physics) undergraduates' and advanced graduate students' understanding of the scientific research process was assessed using two clinical interviews: the "Nature of Science Interview" (International Journal of Science Education 11 (1989) 514; Cognition and Instruction 18 (2000) 349) and the "Nature Nurture Interview" (Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1992) which offered a context (a fictitious conflict of a nature and a nurture theorist about the causes of nervousness in dogs). The participants' answers were then subjected to a categorical analysis using a coding scheme that progresses from an undifferentiated understanding of science as action (Level 1) to an understanding of science as an iterative process of theory formation, testing, and revision with increasing depth of explanation (Level 5). Results indicate that neither group had a clear understanding of the necessity of framework theories for the scientific research process. While no clear developmental progression was found, differences between students majoring in different fields emerged.

☆The research presented in this paper was supported by grant # So 213/10-1 from the Deutsche Forschungsgemeinschaft to Beate Sodian while she was at the University of Wuerzburg, Germany. The study was completed as part of the requirements for a Master's degree in Psychology by Claudia Thoermer.

*Corresponding author.

E-mail addresses: thoermer@edupsy.uni-muenchen.de (C. Thoermer), sodian@edupsy.uni-muenchen.de (B. Sodian).

PII: S0732-118X(02)00009-0

1. Introduction

Students' epistemological beliefs are considered to be an important determinant of academic performance. Since the 1960s, college students' beliefs about the nature of knowledge and the process of knowledge acquisition have been the subject of systematic research (see Hofer & Pintrich, 1997 for a review). Perry (1970), in a longitudinal study of college students' interpretation of their own educational experiences, described a stage-like sequence of conceptions of knowledge, from an absolutist (knowledge is certain and gained from authorities) through a multiplistic (one opinion is as good as another), and a relativistic epistemological stance (knowledge is context dependent), towards an understanding of rational argument, choice and commitment in adulthood (see Baxter Magolda, 1987, 1992; Belenky, Clinchy, Goldberger, & Tarule, 1986; Reich, Oser, & Valentin, 1994, for further research along these lines). Similar descriptions of a developmental sequence of epistemological positions were obtained in studies of reflective judgment (King & Kitchener, 1994; Kitchener & King, 1981; Kitchener, King, Wood, & Davison, 1989), and skills of argument (Kuhn, 1991; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995): Young adolescents typically fail to differentiate between beliefs and facts, either failing to recognize differences in opinion or attributing these to simple factual error. In late adolescence, subjects show an increasing awareness of genuine differences in beliefs which leads to a period of radical relativism (Chandler, Boyes, & Ball, 1990). In adulthood, some people reach a mature epistemological position, recognizing the relativity of interpretation while maintaining canons of rational justification of belief (see Kuhn, Cheney, & Weinstock (2000), for a recent differentiation of such a stage model across different judgment domains). A different line of research has addressed epistemological beliefs without proposing coherent stage-like developmental structures (Schommer, 1990, 1994). The developmental findings emerging from these different lines of research on epistemological beliefs can be characterized in terms of a transition from a "knowledge unproblematic" to a "knowledge problematic" epistemological stance (Carey & Smith, 1993).

Naive epistemological beliefs have long been identified as a major impediment to the achievement of conceptual change in science education (Carey, Evans, Honda, Jay, & Unger, 1989; Carey & Smith, 1993; Cocking, Mestre, & Brown, 2000; Hammer, 1994; Hodson, 1988; Nadeau & Desautels, 1984; Schommer, 1993; Smith, Maclin, Houghton, & Hennessey, 2000; Songer & Linn, 1991). Students who adopt an absolutist epistemological stance will have difficulty in understanding the relation between theories and evidence, and will consequently fail to restructure knowledge in a way that allows them to revise their theories and to engage in the process of scientific inquiry. Lederman (1992) concludes from a review of the literature on students' and teachers' understanding of the nature of science that both students' and teachers' views reflect misconceptions about the nature of scientific knowledge such as the myth of absolute truth and a naive belief that science brings us gradually nearer to the truth by a process of accumulation of facts. Research on disciplinary differences suggests that college students may be more inclined to recognize

divergence of interpretations in the social sciences than in science or engineering where they tend to maintain a belief in the certainty of an unchanging body of knowledge and the authority of experts (Hofer, 2000; Jehng, Johnson, & Anderson, 1993; King & Kitchener, 1994).

To provide a comprehensive in-depth assessment of students' understanding of the nature and purpose of science, the main elements of the process of scientific inquiry, including theoretical ideas, experiments, and results, and the relation among these elements, Carey et al. (1989) developed a clinical interview, probing for students' conceptions of scientific knowledge as well as for their understanding of key concepts like discovery, explanation, and proof. They distinguished between three levels of understanding science[1]: At the least sophisticated level 1, subjects do not differentiate between ideas and facts and see science in terms of activities, such as the production of effects or in terms of accumulation of facts; on level 2, subjects show an understanding of the notions of "testing ideas" and "providing explanations for natural phenomena". The most sophisticated level 3 is characterized by an appreciation of the cyclic, cumulative nature of science with the goal to arrive at an ever deeper understanding of the mechanisms underlying real-world phenomena. Carey et al.'s (1989) study of junior high school students showed an initial understanding of science in terms of concrete activities and collection of facts, and a lack of understanding of the role of ideas (i.e. a predominance of level 1 answers). Through a curricular intervention intended to emphasize theory building and reflection of the theory-building and revision process, these students' understanding could be enhanced by about half a level, that is, they moved from an understanding of science as mere activity to some understanding of science as a process of formulating, testing, and revising ideas. Smith et al. (2000) were able to demonstrate substantial differences between 6th graders who were taught according to a standard science curriculum and predominantly displayed a knowledge-unproblematic, effect based (objectivist) understanding of science, and 6th graders who had experienced a consistently constructivist science classroom and showed a much higher degree of awareness of the role of theory-building, testing, interpretation, and revision in science. That is, even elementary school students are capable of developing a rather sophisticated understanding of science, usually not found until late adolescence or early adulthood if they are exposed to appropriate instruction and classroom experience.

In the present study, we apply the approach taken by Carey et al. (1989), adapted for the older age group, to the study of college students' understanding of the nature of scientific knowledge. We attempt to trace developmental differences in the understanding of the nature of science from late adolescence to adulthood, and to investigate effects of learning and expertise over the college years. Therefore, the present sample included first-year university students and Ph.D. students, all majoring in either physics, chemistry, or biology. Based on the literature on general

[1] As most of the work on metatheoretical understanding of scientific reasoning, their scheme draws from Kuhn's (1962) prominent framework of the philosophy of science, is based on cases in the history of physics.

epistemological development (King, Kitchener, Davison, Parker, & Wood, 1983; Kitchener & King, 1981; Kuhn, Amsel, & O'Loughlin, 1988; Kuhn et al., 2000; Oser & Reich, 1987; Perry, 1970) it was expected that while first-year undergraduates would still hold objectivist positions, the more advanced students, who had been actively involved in a research process for several years, would have arrived at a critical-constructivist understanding of science.

While traditionally more research has dealt with the acquisition of expertise in content domains (e.g., physics, (Chi, Feltovich, & Glaser, 1981; diSessa, 1993)), only recently have more researchers addressed science students' epistemological beliefs. Hammer (1994) interviewed introductory physics students with respect to their beliefs about the structure and content of physics knowledge, and found within-subject consistency over several measurement points as well as between subject differences. Paulsen and Wells (1998) found disciplinary differences in epistemological beliefs with "knowledge unproblematic" epistemological positions occurring more frequently in applied fields, and "knowledge problematic" positions being more frequent in the "pure" sciences and in the "soft" fields. While more studies have dealt with high-school-students' conceptualization of the nature of scientific knowledge (for reviews see Hofer & Pintrich, 1997; Schommer-Aikins, 2002), and older subjects' general intuitive epistemologies, few studies have been conducted addressing the question of whether, and if so, how science university-level-students' initial *epistemologies of science* (rather than general epistemological stance) undergo a process of restructuring over the course of their university education.

To assess students' understanding of the nature of science, we used two interviews, (1) the "Nature of Science" interview (NOS) developed by Carey et al. (1989) probing for an abstract definitional understanding of key elements of the process of scientific inquiry, and (2) an interview providing subjects with contextual support by leading them through a concrete example of a controversy between two scientists (a nature and a nurture theorist) over the causes for "nervousness in dogs" (NNI), during which subjects are asked to comment on each scientist's ideas, experiment, results, and their interpretation of their own and the competing scientist's data. This latter interview was originally designed by Sodian, Carey, Grosslight, and Smith (1992) in order to avoid a possible underestimation of children's and young adolescents' understanding by tools like the NOS. Results show that in fact 12- to 15-yr-olds scored somewhat higher here than in NOS, with fewer subjects remaining on purely effect based answers (no difference between ideas because both scientists are helping dogs). However, even older adolescents and adults (humanities students) mostly did not show a mature understanding of the relation between theories and evidence by construing the conflict of interpretation presented in the interview as a theoretical controversy over one and the same set of data, but either continued to construe theoretical controversy as factual error or recognized the possibility that multiple causes could be obtained without relating causal explanations to the data at hand (Sodian, 1995; Sodian, Thoermer, Schrempp, & Bullock, 1999). In the present study, we investigate whether first year science majors and science graduate students have a more mature understanding of interpretation based on the underlying broader theoretical framework within which a scientist works and thinks.

In summary, the present study was designed with three questions in mind:

(1) How do science students understand the role of theoretical frameworks in the scientific research process?
(2) How does such an understanding develop with increasing scientific practice and exposition to competing models and theories as during graduate studies?
(3) How does performance in an interview tapping abstract-definitional knowledge about components of the scientific inquiry process (as used by Carey et al., 1989) compare to the level of reasoning in a context-enriched format (such as the interview developed by Sodian et al., 1992)?

2. Method

2.1. Subjects

A total of 39 subjects—all students of natural science (biology, chemistry, physics) participated in the study: 18 first-year undergraduates[2] between 19 and 21 yr of age (6 female, 12 male) and 21 graduate students[3] aged between 27 and 32 yr (11 female, 10 male). All undergraduates were interviewed within the first month of their first term. Consequently, none had attended any university-level course on research methods or comparable instruction at that point. Graduate students were in their 2nd or 3rd year of graduate studies. All participants were enrolled at the same Technical University at the time of the study. Participation in the study was voluntary, participants received a small amount of money for participation. Table 1 shows the distribution across subjects for both groups.

2.2. Measures

2.2.1. "Nature of Science Interview (NOS)" (Carey et al., 1989)

Appendix A gives the complete set of questions, those questions reported here are underlined.

The "Nature of Science Interview" assesses subjects' understanding of central concepts of science, such as Hypotheses, Theory, Experiment and their interrelations in an abstract, not context-embedded way. In general, questions were asked in the

[2] In order to be admitted to university level studies in Germany, students have to complete 13 yr of school, starting from age six. The last 2 yr of high school are roughly equivalent to the first 2 yr of college in the US. After this, i.e. when they are approximately 19–21 yr of age, students complete the "Abitur" which entitles them to enter university. They then have to major in one field of study from the first year on. This is the point at which the "undergraduate" sample of the present study was recruited.

[3] Graduate (Ph.D.) students have completed approximately 5 yr of university level studies before entering graduate school. They also have completed a degree equivalent to a Master of Science, including a Master's thesis. Consequently, they are roughly comparable to a third or fourth year graduate student in the US.

Table 1
Distribution of sample by status of expertise and fields of study

	Biology	Chemistry	Physics	Total
Undergraduate	5	6	7	18
Graduate (Ph.D.)	5	10	6	21
Total	10	16	13	39

order in which they appear in the guidelines. However, if a subject's answer seemed ambiguous, further specification was asked for, and if the subject's answer referred to a later question this question was not asked again. If a participant negated a question (such as "Do scientists do experiments?"—"No.") follow-up questions were dropped. Even though all questions of the interview were asked, only a subgroup of questions—those concerned with the understanding of conflicts of data interpretation and alternative theories (Q 4.1, 4.4, 4.5, 4.6) will be analyzed and reported for the present purposes.

2.2.2. "Nature Nurture Interview (NNI)" (Sodian et al., 1992; Sodian, 1995)

Appendix B gives an outline of the entire interview, those questions reported here are underlined.

The "Nature Nurture Interview" assesses subjects' understanding of central concepts and processes of scientific research using a context-enriched format. A special focus is on the differentiation of simple beliefs and interpretive frameworks, an understanding of indirect argumentation from evidence in the justification of beliefs. The interview introduces the phenomenon of "nervous dogs" as well as two fictitious scientists, a nature and a nurture theorist, who treat nervousness according to their theory of etiology (i.e. the nature theorist thinks the cause is a genetically caused hormonal imbalance which can be compensated by adding a missing hormone to the dogs' diet, while the nurture theorist thinks dogs become nervous as a consequence to mistreatment by their owners). For each scientist a highly successful intervention study (nature theorist: adding medicine to diet by vet and owners; nurture theorist: owner training) is described, which is followed by a set of reports in which each scientist summarizes his theoretical point of view, the results and interpretation of his own study and reinterprets the other scientist's findings. It is this part of the interview that will be the focus of attention for the present analyses. Subjects are asked about their understanding of interpretation and reinterpretation, their understanding of the incompatibility of the two competing accounts of the phenomenon (Q 1–4).

The order of the interviews was counterbalanced, i.e. approximately half the subjects in each group received the Nature of Science interview first, and the other half the Nature Nurture Interview. In a later session, subjects also received an additional interview that will not be reported here.

2.3. Procedure

All subjects were interviewed by the same female interviewer (the first author) in a one-on-one situation. Graduate students were interviewed either in an office at the psychology department or in their own office at the university, undergraduates were interviewed either at the psychology department or at their homes.

The interviewer used a written interview guideline throughout the session, and for the context-rich interview, black-and-white illustrations were shown to the participants. The entire session was recorded on conventional tapes with a small tape recorder, which the subjects had agreed to prior to the session. No written notes were taken during the interviews.

2.3.1. Coding

All interviews were recorded on tape and then transcribed. They were then subjected to coding along a qualitative coding scheme based on a critical-constructivist view of science (see next section). For each question the highest level was coded. Over 90% (34 of 37 Nature of Science, 36 of 38 Nature Nurture) of the interview transcripts were coded by two independent coders who were blind to the status of expertise or field of study of the participants. For the Nature of Science Interview intercoder agreement was 80%, for the Nature Nurture Interview 90%.

2.3.2. Coding scheme

The coding scheme used here followed those of Sodian et al. (1992), Carey et al. (1989) and Smith et al. (2000) which reflect three broad categories of epistemic stances commonly described in the literature: naive realism, marked by a lack of understanding of the scientist or learner as a constructor of knowledge (Levels 1 and 2), "relativism", characterized by an appreciation of knowledge as tentative, the result of subjective construction, but a lack of understanding of the testability and rationality of beliefs, and "critical relativism", encompassing both the constructive role of the learner and the testability of beliefs (Carey et al., 1989; Chandler et al., 1990; Kitchener & King, 1981). In order to be able to also code transitions from one stance to the next, we differentiated five levels[4]. Below, these levels will be broadly described and some examples of answers[5] will be provided.

Level 1 answers indicate an *action-understanding of science*, failing to reflect a clear differentiation between thought and activity (i.e. theory and data) or lack of justification (yes/no).

A subject coded on this level for question 1a of the NNI would for example have said: "The result is the same for both scientists: the dogs got calmer and more friendly", focusing entirely on the effect of the treatment, not attending to causal

[4]Our present coding proceeds along the lines of Carey et al.'s (1989) broad scheme. However, we consider it more straightforward to replace her three-level scheme with sublevels (1a, 1b, 1.5, 2, 2.5, 3) by a five-level scheme which considers each sublevel to be a genuine level in itself, but unifies levels 2.5 and 3, as they rarely are found empirically.

[5]All interviews were conducted in German. Example answers are translated by the authors.

mechanisms at all. A Level 1 answer to Q4.1 of the NOS would also focus on effects, such as "When I have an idea I want to try it, and if it doesn't work I'll try until it does." Again, "idea" refers to an effect, rather than to an unobservable causal mechanism, and no notion of testability is detectable.

On *Level* 2, a differentiation between thought and activity is present in that the possibility of having different points of view is acknowledged. However, such differences are traced back to differences in information. Thus, Level 2 comprises both, *naive realist* answers (assuming a knowledge unproblematic access to information) and *radically relativistic* answers (seeing no need to test for compliance with truth). Consequently, answers on this level see conflicts of interpretation either in personal preferences ("He just wants to be right and say something different.") or different access to information ("Maybe they tested two different kinds of dogs."). It should be noted, however, that radical relativist answers of the form that beliefs are generally unjustifiable were hardly observed in the present sample, more often, personal shortcomings, such as stubbornness, were given as reasons for different interpretations at this level.

Answers coded as Level 3 reflect a transitional stage in that they indicate some awareness of diversity of interpretation *but focus on either theory or interpretation*, *failing to integrate both aspects* or show an understanding of the interplay between theory, evidence, and interpretation. However, the "ideal" scientist is seen as not being influenced by his theoretical background, such as in this answer: (Q: Does the theory a scientist holds influence how he interprets experimental findings?—A: "Of course, it shouldn't. Because that would be a real danger. If you always just want to see the theory you are proposing confirmed, you always only look at it from the point of view of your theory. So, I try not to do this. But of course, I'm happy when I see that it fits. I don't want to say that my point of view isn't influenced by my theory. So, I always take a critical look in retrospect and if I see that my experiment does not fit my theory, I have to work some more."). An answer at this level of a subject focusing on theoretical frameworks would be "With his theory he has a view of how things happen. And when there are two scientists with different theories, they have different views. Each one interprets what happens in a different way."

Level 4 reflects a beginning appreciation of critical relativist stance in that the necessary *role of theoretical frameworks for the interpretation of evidence* is acknowledged at least implicitly. Also, a beginning understanding of the cyclic nature of the scientific research process is evident; however, a clear differentiation between theories and hypotheses is not present, as in the following example: "Yes, when you insist on the theory, like this theory has to be right.... Sure, you need the theory as a basis for your work in so far it does have an influence, it's just when you get different results than what would support the theory, you have to be able to deviate from it."

Level 5 exceeds Level 4 in that *theory dependence of interpretation and the conflict of interpretations stemming from competing theories* are explicitly acknowledged. Subjects at this level also propose a contrastive test to decide between conflicting interpretations. Thus, answers at this level demonstrate a differentiation of theoretical frameworks, hypotheses and experiments, with a clear notion of

testability. An example would be: (Q: How can you decide between different interpretations?—A: "Well, that often is not so simple at all. Sometimes you find an experiment that excludes the alternative interpretation, and if this is not possible, which is also quite frequently the case, you can't decide at that point. Until someone comes up with a new finding and postulates a new theory that doesn't contradict the previous stuff and everybody can agree on that.").

3. Results

Results will first be reported by comparing overall means with parametric methods, followed by nonparametric explorations of overall levels. In the final section, results for single questions and comparisons across questions and interviews will be presented.

3.1. Overall mean scores

Table 2 shows mean scores for NNI/NOS by group and major field of study. A mixed design 2 (interview) $\times$ 2 (group) $\times$ 3 (field of study) $\times$ 2 (order) $\times$ 2 (sex) ANOVA showed a significant effect of interview ($F(1,15) = 8.63, p = 0.010$). Subsequent analyses will collapse data across order and sex. Separate one-way ANOVAS for each of the interviews showed effects of major fields of study for both interviews (NNI: $F(2,35) = 6.46, p = 0.004$; NOS: $F(2,34) = 4.97, p = 0.013$), with the physics students outperforming both the students of chemistry (Tukey $p = 0.006$) and biology (Tukey $p = 0.022$) in NNI, and those of biology in NOS (Tukey $p = 0.011$). For NOS the difference between physics and chemistry students did not reach significance (Tukey $p = 0.110$). Also, for NOS (but not NNI) a nonsignificant effect of group (first-year vs. graduate) ($F(1,35) = 2.58, p = 0.117$), with graduates arguing at a higher mean level than first-year students, was found. Performance between the interviews differed significantly for the undergraduates, who achieved higher in the NNI than the NOS ($t(17) = 3.21, p = 0.005$), while this difference was not found to reach significance for the more advanced Ph.D. students ($t(17) = 1.45, p = 0.164$). Across groups, the effect of interview (again, as always, showing higher scores for NNI than NOS) was most pronounced for the biology majors ($t(9) = 3.08, p = 0.013$), and did not reach significance for the chemistry ($t(13) = 1.42, p = 0.178$) or physics ($t(11) = 1.22$, n.s.) students.

Table 2
Mean scores for NNI/NOS by group and major fields of study

	Biology	Chemistry	Physic	Overall
Undergraduates	3.05/2.2	2.92/2.29	3.52/3.07	3.19/2.57
Ph.D. student	3.2/2.35	3.19/3.0	3.92/3.85	3.41/3.05
Overall	3.12/2.27	3.08/2.72	3.71/3.4	3.32/2.82

Table 3
Overall levels (%) (dichotomized) by interview

	NOS-OL1	NOS-OL 2	All
NNI-OL 1	17	8	25
NNI-OL 2	44	31	75
All	61	39	100

Taken together, the effect of subject was more pronounced than the effect of expertise (i.e. student status), with the physics students at both levels showing more sophistication than students of either of the other two subjects. Also, the NOS proved to be more "difficult" for subjects, even though less so for the physics than biology or chemistry majors.

3.2. Overall analysis by levels

As the levels subjects' answers were assigned to vary considerably even within interview and subject, it is desirable to also include analyses reflecting the consistency of answers, rather than just mean values. Therefore for each interview, additional overall levels were calculated. As answers on levels 1 and 2 show clearly less sophistication (no understanding of the role of mental activities such as thinking and inference in the scientific research process) than the others, subjects were assigned an overall level (OL) of 1 if they answered more than one of the four questions per interview at level 2 or lower, and an OL of 2 if they answered none at level 1, maximally one at level 2, and all others at level 3 or higher. Table 3 shows the percentages of subjects who were assigned OL of 1 and 2 for each interview. As can be seen from Table 3, clearly more subjects were assigned OL 2 for the NNI than the NOS (McNemar test, $N = 36, p = 0.004$). Separate analyses showed that this effect was significant for the undergraduates (McNemar, $N = 18, p = 0.039$), but not quite for the Ph.D. students (McNemar, $N = 18, p = 0.109$). Twenty-eight percent of the undergraduates and 33% of the graduate students achieved OL 2 in both interviews. Forty-four percent of the undergraduates and 44% of the Ph.D. students were assigned OL 2 for the NNI but not the NOS, while the reverse was true only for 6% of the undergraduates and 12% of the graduate students.

Due to small Ns the effect was not significant for any single major field of study. While none of the biologists was assigned OL 2 for both interviews, the majority (62%) was assigned OL 2 for the NNI, but not the NOS, only 12% showed the reverse pattern. The chemistry majors displayed a similar pattern with 25% reaching OL 2 for both interviews, but still more (44%) only for the NNI, but not for the NOS, and only 12% for the NOS, but not the NNI. In contrast, among the physicists, the majority (58%) reached OL 2 for both measures, 33% had an OL 2 for the NNI only and none showed the reverse pattern.

Overall, the results of the analyses taking into consideration the overall levels confirm those found by parametric methods.

3.3. Overall scores compared to Carey et al. (1989)

In order to be better able to compare the performance of subjects in our study to Carey et al.'s junior high school students, we collapsed parts of our category system. For that purpose, Levels 1 and 2 were considered equivalent to Carey et al.'s level 1 (science as activity), Levels 3 and 4 as equivalent to Carey et al.'s level 2 (science for explanation, simple access to knowledge), and Level 5 as equivalent to level 3 of Carey et al. (science as search for deeper explanations; cyclic character, indirect testability of contentions). When scores were transformed in this way, the vast majority (73%) of subjects had a mean level of 2 (according to Carey et al.) for the NNI, while performance was more mixed for the NOS (level 1: 35%, level 2: 54%, level 3: 11%). As in our five-level coding system, even if recoded according to Carey et al.'s three levels, there was an effect of interview (Wilcoxon-Test $N = 36, z = 2.83, p = 0.005$), as only eight subjects achieved a higher mean score on the NOS, six received the same score, while 22 scored higher on the NNI than the NOS. While separate one-way ANOVAS showed no difference between groups for either interview, a significant effect of subject for both interviews (NNI: $F(2, 25) = 5.25, p = 0.01$; NOS $F(2, 34) = 4.61, p = 0.017$) emerged: the physics students again reached higher scores than those of biology and chemistry.

Compared to Carey et al.'s (1989) junior high school students' performance on the NOS, who reached a mean level of 1.0 before and 1.55 after a training intervention, some developmental progression through adolescence and young adulthood can be seen with performance on the NOS at a mean level of 1.67 and 1.97 on the NNI.

3.4. Question-based analyses

3.4.1. Nature nurture interview

As can be seen in Figs. 1 and 2, answers were not on consistent levels across the questions of the interview, even though all of them deal with an understanding of the reasons, role, and possibilities to deal with differing interpretations of data.

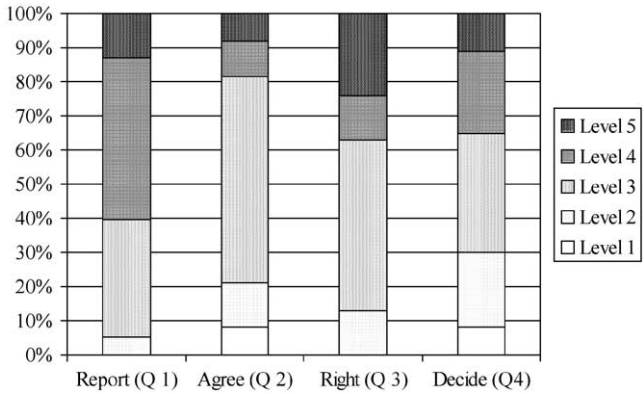

Fig. 1. Distribution of subjects' answers across levels for questions of the NNI.

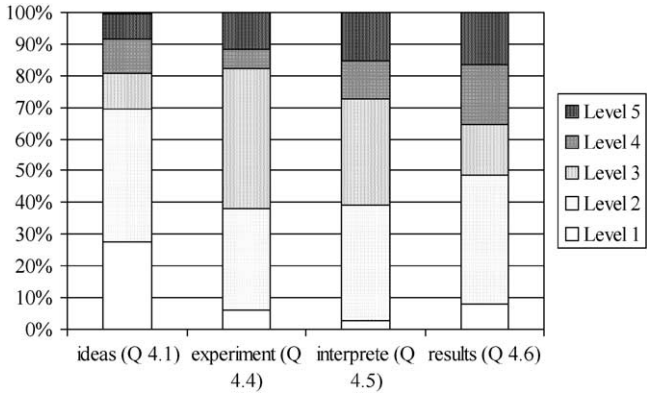

Fig. 2. Distribution of subjects' answers across levels for questions of the NOS.

Question 1 While a majority (86%) of subjects[6] produced answers coded on Level 4 when asked about important differences between the scientists' reports, thereby demonstrating that they were able to distinguish data and differing interpretations, especially Questions 2 and 3 proved to be harder for them. There also was a tendency for subjects who had received the NOS first to answer at higher levels (Fisher's exact test $N = 38, p = 0.65$). All five subjects whose answers were coded at Level 5 (i.e. who spontaneously produced the difference in interpretation and reinterpretation as the main difference between the reports and gave an account connecting data interpretation to the different underlying framework theories) had received the NOS first. Also, while even for undergraduates Level 4 was the most frequent level of answering this question for both graduates and undergraduates, no undergraduate answer was coded at Level 5.

The most frequent Level of answering *Question* 2 concerning whether the two scientists agree on the meaning of the respective pattern of results was only Level 3, that is, even though subjects were able to give both interpretations and reinterpretation they did not comparatively comment on this in a way indicating an implicit understanding of the theoretical basis for these differing interpretations. Most frequently, they focused only on one scientist, but failed to contrast this view to that of the other researcher (e.g. "He thinks that medical treatment will cause that the dogs are treated better, which confirms his theory"). Again, no effect of expertise status (group) is present, but physicists' answers tended to be coded at higher levels (Fisher's exact test $N = 38, p = 0.05$).

Question 3 offered a number of prompts probing for subjects' understanding that both theories offered mutually exclusive causal chains (born with genetic deficit vs. behaviorally acquired after birth), asking whether both could be right at the same time. While only few subjects mentioned this problem at first, when prompted by

[6] Due to uninterpretable answers or interviewer errors data of some subjects for different single questions are missing. Percentages are given relative to the number of codable subjects.

"at the same time" most denied this possibility, giving reasons mostly coded at Level 3. They either realized the different causal chains or that each scientist had a different underlying theory, or that each scientist had some evidence to support his view; however, they failed to interconnect these aspects. However, few subjects followed the strategy found in younger and less sophisticated samples more frequently, namely, the construction of different groups (i.e. scientists by massive coincidence have sampled different kinds of dogs and each found the right answer for his sample) in order to circumvent this conflict. Still, anecdotal evidence indicates that even the present science students were challenged by this question, indicated by a lot of confirmatory questions about past facts of the story, relatively long pauses and changes of focus within answering the question.

Question 4 asked about possible measures to decide between both accounts. While the median was Level 3, suggesting further elaboration of or experiments supporting one theory, again physicists tended to answer at higher levels than the other students (Kruskal–Wallis Test $\chi^2_{(2,N=37)} = 5.93, p = 0.052$). Only 13 subjects (4 undergraduates, 8 graduates) gave answers codable at Level 4 or 5, explicitly demanding a critical experiment in order to exclude one causal chain or suggesting a modification of the causal assumptions in order to be able to merge them into a coherent theory.

3.4.2. Nature of science interview

*Q*4.1: *How do the ideas or theories a scientist has influence the experiments he or she does*?

Forty-two percent of all participants saw some relation between theory and experiment, however, they neither considered this conducive nor necessary, but rather considered it an unfortunate bias (Level 2). Twenty-eight percent even gave answers coded on Level 1, saying that "if you have an idea, that's just how you do it", showing no understanding of experimentation as a means to subject hypotheses to a critical test. A total of 11% saw an experiment as a means of testing a simple idea or guess; however, indicated no awareness of the necessity of theory-driven experimentation. Only one-fifth of the sample implicitly (11%) or explicitly (8%) demonstrated an understanding of experimentation as a means of testing hypotheses derived from a larger theoretical framework, and of the relevance of experimental results for the framework theory. Of the seven subjects whose answers were coded on Levels 4 and 5, four were physicists. Overall, this question yielded answers on lower levels than the other questions of the NOS. There was a marginally significant effect of subject (Fisher's exact test $N = 36, p = 0.015$), with physicists outperforming the two other subject groups.

*Q*4.4: *Does the theory a scientist has influence his interpretation of experimental results*?

Subjects performed slightly better on this question, with the median on Level 3 (44%). Answers at Level 3 typically conceded that theoretical views would influence the interpretation of experimental results; however, they did not elaborate on this or considered it as nondesirable (as opposed to Level 2 (32%) where this was seen as an arbitrary act, Level 3 included the notion of the necessity of theory-driven research). A typical example of a Level 3 answer would be: "Of course, it shouldn't. That

would be a real danger. If you always just want the theory you have confirmed, you always just look at it (the experiment) from the point of view of your theory. So I try not to do this, though of course I'm happy if it fits. I don't want to exclude that my point of view is influenced by my theory. But I always take another critical look afterwards and when my experiment does not confirm my theory I know I have to do some more." Only 18% acknowledged the function of testing hypotheses derived from a theory and the consequences experimental results have for the truth value of a theory implicitly or explicitly.

*Q*4.5: *Is it possible that the same experimental results are interpreted differently by different scientists*?

As in the previous NOS questions, the majority of answers was coded at Levels 2 (36%) and 3 (33%), with subjects realizing the possibility of divergent interpretations, but either judging them as an act of arbitrariness or even fraud (Level 2) or of faulty scientific practice (Level 3). Only 5 subjects (15%) showed an understanding of the possibility of divergent interpretations of experimental results derived from competing framework theories.

*Q*4.6: *How can it be decided in such a situation which scientist or theory is right*?

This question was most frequently answered on Level 2 (40%), that is those subjects asked for "more experimental evidence"; however, without specifying the kind of experiment or evidence required. Six (16%) subjects coded on Level 3 asked for "more experiments in order to further investigate one of the ideas". A total of 13 subjects (35%) were coded on Levels 4 and 5, explicitly demanding a decisive experiment, contrasting both views ("Each scientist has to say what the critical factors in his models are, where they differ. Then these factors are manipulated and you have to see what the results is."), of which a subgroup of seven subjects (four of them physics Ph.D. students) required a refinement and integration of competing theories in case the results of contrastive experiments would not clearly support one or the other theory.

4. Discussion

How do young scientists construe the nature of scientific knowledge? Previous research on children's and adolescents' epistemologies of science has indicated a developmental progression from a misconstrual of science in terms of activities or collection of facts in children towards an implicit appreciation of ideas or theories in adolescents or young adults (Carey et al. 1989; Sodian et al., 1999). Research on the effects of science teaching indicates that even in elementary school children an (implicitly) constructivist epistemology of science can be achieved by a long-term explicitly constructivist science curriculum (Smith et al., 2000). One aim of the present study was to investigate developmental and expertise-related progress in German university students' epistemologies of science: Do science students at university level articulate an explicit understanding of the process of theory construction and revision in science? Do they apply such a constructivist epistemology of science to a concrete example of scientific controversy outside their

own field of expertise? How do undergraduates differ from graduate students with respect to their epistemologies of science?

The present results of two interview studies conducted with first-year undergraduates and graduate students in physics, chemistry, and biology indicate that an appreciation of the role of interpretive frameworks in the construction of scientific knowledge remains at best implicit even in graduate students who are actively involved in an empirical research program. Answers at Levels 4 and 5, characterized by a more explicit articulation of the role of theories or interpretive frameworks in the construction of scientific knowledge, were extremely rare. This is also reflected in the mean scores, with only the physicists exceeding a mean score of 3 for both interviews.

Perhaps the most surprising result of the present study was the absence of measurable differences between undergradutes' and graduate students' epistemologies of science. While novices differ dramatically from experts in their conceptualization of their respective fields of study (Chi et al., 1981), their domain-general metaconceptual understanding of science does not seem to undergo a similar restructuring process. It should be mentioned that unlike Chi et al. (1981) we did not compare undergraduates with physics professors, nor did we contrast a mixed sample of undergraduates with Psychology faculty, like Schunn and Anderson (1999), but compared purely science samples of undergraduate and graduate students, that is, the difference in level of expertise was not as large as in other studies. However, we can safely assume that our subjects' *content knowledge* underwent considerable restructuring over the course of a 5–7-yr university education. Apparently, it is possible to successfully complete a university education in a natural science, without developing an explicitly articulated metaconceptual notion of interpretive frameworks or theories. Our subjects clearly articulated an understanding of testing, and a notion of explanation, but they tended to neglect the necessity of framework theories, and to see competing local theories as a symptom of faulty scientific practice. Although our subjects were able to recognize theories, and to *work within* competing frameworks in the NNI Interview, they tended to hold an objectivist view of science as their *normative ideal*. When questioned informally about their own work, some of our graduate students expressed disappointment with the situation of being restricted by their advisor's approach, rather than being able to "interpret results objectively". It is possible that they were discouraged from developing a critical constructivist epistemology of science by being kept in an intellectually dependent role[7] throughout their graduate career (possibly linked to the fact that most of the sample were paid by their advisors' grants, though none of them directly industry-sponsored).

Only slight increases in overall scores were found in comparison with the grade seven students whom Carey et al. (1989) interviewed with their first version of the NOS. Even though our sample of university students did not display an explicit

[7] The German university system is sometimes accused of being "authoritarian" and restraining the intellectual growth of its students, but we know of no systematic research addressing this issue in cross-cultural comparison.

notion of theories as interpretive frameworks, the development from Carey et al.'s level 1/1.5 to level 2 must be considered a major advancement since it marks the progression from a naïve-realist view of science to a critical-constructivist understanding of the scientific research enterprise with an articulated understanding of critical testing and explanation.

With regard to the influence of the interview format, the hypothesized positive effect of providing contextual support was confirmed, as indicated by the higher overall scores achieved in the NNI than the NOS. The differences found between the two interviews in the present sample parallel the ones found in younger age groups indicating a facilitating effect of contextual support. However, this effect was stronger for the undergraduate students, indicating that the Ph.D. students were more apt at verbalizing their understanding.

An unexpected, but relatively strong effect was that of field of study. How can this be explained? Why did even the physics undergraduates display a more explicitly constructivist understanding of science than did the graduate students of the other faculties? Why did even the physics undergraduates answer the abstract NOS questions at a more sophisticated level than the biology and chemistry Ph.D. students? One speculation would be that this reflects the general state of the respective fields of study with regard to the Kuhnian scheme of "normal science" vs. states of "revolution". The most popular examples for groundbreaking conceptual changes do come from the domain of physics, which might leave such issues more up front in this field. In contrast, biology and chemistry are currently mainly making news by "discovery" of new elements, new measurement techniques or the like, while the conceptual developments underlying these developments usually do not make the headlines. Furthermore, in the German system, students take some extended courses in their future field of study throughout their last 3 yr of high school (comparable to the first years of college), so even prior to entering university these students might have had different epistemological backgrounds.

Studies of scientists' thinking in their own research practice (Dunbar, 1995; Dunbar & Klahr, 1989) have shown that mature scientists pay a great deal of attention to unexpected results, and continuously make sure that their current goals do not hinder them from considering alternative theories or alternative ways of conducting experiments. It appears that such scientific practice requires an explicit notion of the role of theories in the construction of scientific knowledge. This notion was not found to be solidly in place for any of the groups who participated in this study, and metaconceptual understanding varied considerably between different disciplines even though all of them are assumed to rely on grossly the same heuristics in theory building and testing (namely that of natural science).

It appears that future research on scientists' epistemologies of science should address the relation between their epistemological ideas (as articulated in an interview), and their thinking in the lab. Also, the present research should be taken further, by interviewing mature scientists, and comparing their reflections on the process of scientific inquiry with the notions articulated by graduate students.

Acknowledgements

The authors wish to thank Inge Schrempp, Sigrid Weinmann, Verena Vetter, and Eva Stumpf for invaluable help with project coordination, transcription and analysis.

Appendix A. The nature of science interview (after Carey et al., 1989)

---

1. General
   1.1. What do you think science is all about?
   1.2. What is the goal of science?
   1.3. What do scientists do? How do they achieve their goals?
   1.4. Do scientists ask questions? What kinds of questions?
      1.4.1. How do scientists answer their questions? Could you give an example of a scientific question and what a scientist would do to answer it?
2. Experiments
   2.1. What is an experiment?
   2.2. Do scientists do experiments?
      2.2.1. Why do scientists do experiments?
         2.2.1.1. What does an experiment show a scientist about his/her ideas or hypotheses?
      2.2.2. How does a scientist decide what experiment to do?
3. Hypotheses
   3.1. What is a hypothesis?
      Is a hypothesis the same as an educated guess or is there a difference? Of what kind?
4. Theories
   4.1. *Do the ideas or theories a scientist has about something influence the experiments he or she does? How*?
   4.2. How does a scientist get ideas/a theory? (*Theory/idea used depending of which term the subject introduced)
   4.3. What is a theory? Do scientists have theories?
   4.4. *Does the theory or ideas a scientist has influence his or her view of an experiment? How*?
   4.5. *Is it possible that the same results are interpreted differently by different scientists? Why and how*?
      4.5.1. *How can one decide which scientist or which theory is right in such a case*?
   4.6. *If a scientist does an experiment that does not yield the expected results, would* he/she *consider this a "bad" experiment? Why? Can something be learned from this*?
   4.7. Let us say a scientist does an experiment in order to test his ideas. Would he/she do an experiment that shows that his ideas are wrong? Why (not)?
   4.8. What happens to an idea/theory once it has been tested?
   4.9. Do scientists change their ideas? If so, under what circumstances and why?

4.10. Do scientists sometimes change their entire theory? If so, under what circumstances and why?
4.11. Do scientists always achieve their goals? If not, why?
4.12. Do scientists make mistakes or err? How?

Appendix B. The nature nurture interview

Phenomenon: Nervous dogs: bark and bite more than others and have higher pulse, fever

Nurture-theorist (Scientist A): Nervousness caused by wrong treatment by owners.

Q: How could scientist A test his idea?

Scientist A: Nervousness should get better if owners are trained at how to treat dogs. Experiment: Group 1: owner training; Group 2: no training.

Q: What results does scientist A expect? (*Prompt: With respect to barking and biting /pulse and fever for each group?*)

Q: What would be an unexpected result?

Results of scientist A's experiment: Group 1: behavioral and bodily symptoms improve, Group 2: no change.

- part on confounded variables
- part on valid/invalid conclusions

Scientist B (nature-theorist) does a different experiment: Group 1 gets medicine, group 2 gets no medicine.

Q: Do you think he has the same ideas about why some dogs are nervous as scientist A? What are his ideas?

Scientist B thinks that nervous dogs lack hormone from birth on. This causes bodily symptoms, which in turn cause behavioral symptoms.

…Both scientists hear about the other's results and write report.

Report of scientist B: Improvement of experimental group in scientist A's experiment due to better feeding schedule and food quality, which partly compensates for lack of hormone. However, complete recovery can only be achieved by medicine, which contains the hormone that nervous dogs lack from birth on. This leads to normalization of pulse and fever that in turn leads to improvement of nervous behavior.

Report of scientist A: Improvement of experimental group in scientist B's experiment due to owner's visit to vet when medicine is administered. They observe how vet treats dog and then can take better care of their dog, too. However, the real reason for nervousness is wrong treatment. When dogs are treated better, they do not bark and bite so much any more, are less excited, which in turn leads to decrease in pulse and fever.

*Q*1: *Are there any important differences between the two reports? If yes, what differences*?

How come the two reports are so different—both scientists know the results of both experiments?

*Q*2: *Think about the results of the experiment with the medicine. All those dogs, which received medicine, are not nervous anymore. All those dogs who were not given the medicine are still nervous.*

Do the scientists agree on what this results shows or do they disagree?

Scientist B (*nature theorist*) *thinks that his results show that nervousness in dogs can be treated with medicine, because the cause is a missing hormone. The medicine decreases blood pressure and so the dogs become calmer and more friendly, and stop barking and biting. Does scientist A* (*nurture theorist*) *also think that this is what the study with the medicine shows*?

What does scientist A think?

*Q*3: *Could both scientists be right with what they wrote in their reports*?

Scientist A says that dogs become nervous after they are born because of the way their owners treat them. Scientist B says the dogs are born nervous. Could both scientists be right?

Scientist B thinks that those dogs that were given medicine are less nervous now because the missing hormone was replaced. Scientist A also thinks that the dogs in scientist B's experiment are less nervous now. But not because they were given the medicine, but because their owners now treated them better. Could both scientists be right at the same time?

*Q*4: *Could anything be done to decide between the two? If so, what*?

References

Baxter Magolda, M. B. (1987). The affective dimensions of learning: Faculty–student relationships that enhance intellectual development. *College Student Journal*, *21*, 46–58.

Baxter Magolda, M. B. (1992). *Knowing and reasoning in college: Gender related patterns in students' epistemological development*. San Francisco: Jossey Bass.

Belenky, M. F., Clinchy, B. M., Goldberger, N. R., & Tarule, J. M. (1986). *Women's ways of knowing: The development of self, voice, and mind*. New York: Basic Books.

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). "An experiment is when you try it and see if it works": A study of grade seven students' understanding of the construction of scientific knowledge. *International Journal of Science Education*, *11*, 514–529.

Carey, S., & Smith, C. (1993). On understanding the nature of scientific knowledge. *Educational Psychologist*, *28*, 235–251.

Chandler, M. J., Boyes, M., & Ball, L. (1990). Relativism and stations of epistemic doubt. *Journal of Experimental Child Psychology*, *50*, 370–395.

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

Cocking, R. R., Mestre, J. P., & Brown, A. L. (2000). New developments in the science of learning: Using research to help students learn science and mathematics. *Journal of Applied Developmental Psychology*, *21*, 1–11.

DiSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction*, *10*, 105–125.

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg, & J. Davidson (Eds.), *Mechanisms of insight* (pp. 365–395). Cambridge, MA: MIT-Press.

Dunbar, K., & Klahr, D. (1989). Developmental differences in scientific discovery processes. In D. Klahr, & L. Kotovsky (Eds.), *Complex information processing. The impact of Herbert A. Simon. 21. Carnegie Mellon Symposium on Cognition* (pp. 109–143). Hillsdale, NJ: Erlbaum.

Hammer, D. (1994). Epistemological beliefs in introductory physics. *Cognition and Instruction*, *12*, 151–183.

Hodson, D. (1988). Towards a philosphically more valid science curriculum. *Science Education*, *72*, 19–40.

Hofer, B. (2000). Dimensionality and disciplinary differences in personal epistemology. *Contemporary Educational Psychology*, *25*, 378–405.

Hofer, B., & Pintrich, P. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, *67*, 88–140.

Jehng, J.-C., Johnson, S. C., & Anderson, R. C. (1993). Schooling and students' epistemological beliefs about learning. *Contemporary Educational Psychology*, *18*, 23–35.

King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgement: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco: Jossey Bass.

King, P. M., Kitchener, K. S., Davison, M. L., Parker, C. A., & Wood, P. K. (1983). The justification of beliefs in young adults: A longitudinal study. *Human Development*, *26*, 106–116.

Kitchener, K. S., & King, P. M. (1981). Reflective judgement: Concepts of justification and their relationship to age and education. *Journal of Applied Developmental Psychology*, *2*, 89–119.

Kitchener, K. S., King, P. M., Wood, P. K., & Davison, M. L. (1989). Sequentiality and consistency in the development of reflective judgement: A six-year longitudinal study. *Journal of Applied Developmental Psychology*, *10*, 73–95.

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

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. Orlando, FL: Academic Press.

Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development*, *15*, 309–328.

Kuhn, D., Garcia-Mila, Z., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monograph of the Society for Research in Child Development*, *60*, v–128.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of Research in Science Teaching*, *29*, 331–359.

Nadeau, R., & Desautels, J. (1984). *Epistemology and the teaching of science*. Ottawa: Scientific Council of Canada.

Oser, F. K., & Reich, K. H. (1987). The challenge of competing explanations. The development of thinking in terms of complementary "theories". *Human Development*, *30*, 178–186.

Paulsen, M. B., & Wells, C. T. (1998). Domain differences in the epistemological beliefs of college students. *Research in Higher Education*, *39*, 365–384.

Perry Jr., W. G. (1970). *Forms of intellectual and ethical development in the college years*. New York: Academic Press.

Reich, K. H., Oser, F. K., & Valentin, P. (1994). Knowing why I now know better: Children's and youth's explanations of their world view changes. *Journal of Research on Adolescence*, *4*, 151–173.

Schommer, M. (1990). Effect of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology*, *82*, 498–504.

Schommer, M. (1993). Epistemological development and academic performance among secondary students. *Journal of Educational Psychology*, *85*, 406–411.

Schommer, M. (1994). An emerging conceptualization of epistemological beliefs and their role in learning. In R. Garner, & P. A. Alexander (Eds.), *Beliefs about text and instruction with text* (pp. 25–40). Hillsdale, NJ: Erlbaum.

Schommer-Aikins, M. (2002). An evolving theoretical framework for an epistemological belief system. In B. K. Hofer, & P. R. Pintrich (Eds.), *Personal epistemology. The psychology of beliefs about knowledge and knowing* (pp. 103–118). Mahwah, N.J: Lawrence Erlbaum.

Schunn, C. D., & Andersen, J. R. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive Science*, *23*, 337–370.

Smith, C., Maclin, D., Houghton, C., & Hennessey, G. M. (2000). Sixth-grade students' epistemologies of science: The impact of school science experiences on epistemological development. *Cognition and Instruction*, *18*, 349–422.

Sodian, B. (1995). *Children's and lay adults' understanding of the relation between theories and evidence*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Sodian, B., Carey, S., Grosslight, L., & Smith, C. (1992). *Junior high school students' understanding of the nature of scientific knowledge. The notion of interpretive frameworks*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Sodian, B., Thoermer, C., Schrempp, I., & Bullock, M. (1999). *Understanding the nature of scientific knowledge—What develops from early adolescence to adulthood*? Paper presented at the Tenth European Conference of Developmental Psychology, Spetses, Greece, September 1999.

Songer, N. B., & Linn, M. C. (1991). How do students' views of science influence knowledge integration? *Journal of Research in Science Teaching*, *28*, 761–784.