Third Misconceptions Seminar Proceedings (1993)

Paper Title: **WHAT CHILDREN KNOW ABOUT METASCIENCE**
Author: Samarapungavan, Ala

Abstract: This paper will present the results of two experiments that examine certain aspects of children's metascientific knowledge. These experiments examine whether children use metascientific criteria such as conceptual coherence and empirical consistency to evaluate competing explanations for natural phenomena.

Keywords: Concept Formation,Educational Methods,,Cognitive Development,Scientific Concepts,Thinking Skills,Metacognition,,
General School Subject: Chemistry
Specific School Subject: Earth Science
Students: Elementary School

Macintosh File Name: Samarapungavan - Metascience
Release Date: 12-16-1993 C, 11-6-1994 I

Publisher: Misconceptions Trust
Publisher Location: Ithaca, NY
Volume Name: The Proceedings of the Third International Seminar on Misconceptions and Educational Strategies in Science and Mathematics
Publication Year: 1993
Conference Date: August 1-4, 1993
Contact Information (correct as of 12-23-2010):
Web: www.mlrg.org
Email: info@mlrg.org

A Correct Reference Format: Author, Paper Title in The Proceedings of the Third International Seminar on Misconceptions and Educational Strategies in Science and Mathematics, Misconceptions Trust: Ithaca, NY (1993).

your work, consider supporting our ability to support you by purchasing a subscription to the collection or joining the Meaningful Learning Research Group.

----- -----

WHAT CHILDREN KNOW ABOUT METASCIENCE

Dr. Ala Samarapungavan
Department of Educational Studies
Purdue University
U.S.A

INTRODUCTION

This paper will present the results of two experiments that examine certain aspects of children's metascientific knowledge. These experiments examine whether children use metascientific criteria such as conceptual coherence and empirical consistency to evaluate competing explanations for natural phenomena.

As the metaphor of the "child-as-scientist" has gained currency in psychology and education, it has also generated considerable debate about the degree to which the child's reasoning actually resembles that of the scientist. One influential approach to these matters was that of Inhelder and Piaget (1958) who claimed that pre-adolescent children were incapable of scientific reasoning.

The knowledge-based approach to cognitive development (Carey, 1985) emerged in part as a response to Piaget's claims. Recent research has shown that when tested with tasks that are simple and meaningful, children demonstrate an impressive array of reasoning skills such as the ability to reason causally (Brown, 1990), analogically (Vosniadou, 1989), and to make inductive inferences (Gelman & Markman, 1986).

A strong form of the knowledge-based approach to issues of scientific reasoning is the "naive theory" approach (Carey, 1985; McCloskey & Kargon, 1988). This approach ascribes apparent differences in the reasoning of novices and scientists to differences in the conceptual content of their respective "theories" for physical phenomena. For example, Wiser (1988) notes that because novices have different theories for heat and temperature phenomena from those of experts, their reasoning on certain problems involving these phenomena appears illogical in terms of the expert model although it is internally consistent. Implicit in the metaphor of "naive theories" is the notion that lay people in general and children in particular construct conceptual frameworks or theories to understand their world (Carey, 1985; Driver and Easley, 1978). It has been proposed that these naive theories, though quite different in content from their scientific counterparts, are explanatory, causal, and generative in that they lead to predictions about hitherto unobserved phenomena (Brewer &

Samarapungavan, 1991). Examples of such naive theories have been documented among children and adults in a variety of domains such as classical mechanics (McCloskey & Kargon, 1988) and astronomy (Vosniadou and Brewer, 1992). Researchers within this tradition attribute performance differences between lay people and experts to differences in domain knowledge and the phenomena available for consideration rather than to differences in any underlying cognitive competence.

Recently this last assumption has come under considerable criticism from researchers like diSessa (1988) and Kuhn (1989). While the above researchers readily acknowledge that novices spontaneously generate concepts to account for what they experience, they argue that novices do not evaluate the adequacy of their concepts in the same way that scientists do. For example, diSessa (1988), and Solomon (1983) have proposed that novices unlike scientists do not appreciate the need for internal consistency or conceptual coherence in their "theories." Although the lay adult is not considered immune from such metacognitive defects, the child is considered to be particularly susceptible. Recently, Kuhn, Amsel and O'Loughlin (1988) concluded from their empirical work that young children do not fully differentiate between their theories as conceptual entities and the external evidence for these theories. They are therefore particularly prone (in Kuhn et al's view) to errors of confirmatory reasoning and insensitive to disconfirmation. It is also suggested by the above researchers that children and to a lesser extent lay adults lack the skills involved in understanding the meaning of evidence once it is sufficiently differentiated from theory. The research to be described here, was undertaken to further examine what children know (or do not know) about such metascietific aspects of scientific reasoning.

The current research is based upon a pragmatic model of scientific rationality (Laudan et al., 1986) which suggests that rational scientific judgments are based on the compatibility of the meaning of new ideas with existing ones that are thought to be well founded. Past research on children's ability to coordinate theories with evidence has focused on the logic of disconfirmation. However, post-positivist philosophers of science (Kuhn, 1977; Laudan, 1977) deny the primacy of disconfirmatory tests in determining a theory's success or failure. They note that many factors contribute to judgments about the success of theories such as the relative precision of theories in predicting important domain phenomena, and their generative success or ability to account for novel, qualitatively diverse phenomena. For example, modern evolution theory was not established by a logic of disconfirmation but rather built upon pieces of empirical evidence that were consistent with its explanatory mechanisms, such as the fossil record.

A second point made by most of the post-positivist philosophers of science (Laudan, 1977; Thagard, 1991) is that theory production and theory revision are very difficult and time-consuming processes for the scientist. They argue that it is cognitively irrational for a scientist to give up a theory that works well for at least some subset of phenomena in the face of new anomalies unless there is an alternative theory that can give coherence to domain phenomena in its place. Therefore metaconceptual criteria for accepting or rejecting theories such as empirical consistency, conceptual coherence, and parsimony, are more likely to be applied when are well articulated theoretical alternatives to choose from.

One implication of the pragmatic model for psychologists interested in the development of scientific reasoning is that the domain specific ideas of subjects must be taken into account in designing experiments to determine whether they can apply the above mentioned kinds of metaconceptual criteria to theory evaluation and theory selection. Subjects are unlikely to give up an idea they believe to be well founded even in the face of new empirical or conceptual ambiguities if the alternative idea is implausible to them on the basis of its conceptual content.

I undertook an empirical investigation to determine whether or not novices can use certain metaconceptual criteria for theory evaluation and theory selection when the plausibility of conceptual content has been controlled for. The central hypothesis of the research was that children possessed much of metacognitive knowledge for theory evaluation and selection used by scientists. Specifically, it was thought that on theory choice tasks, children would apply important metaconceptual criteria for theory evaluation and selection in all cases where the competing theories were both compatible or both neutral with regard to the children's' domain specific beliefs. In the following section, the two experiments undertaken to investigate what children know about metascience will be described. Since the two experiments use the same methodology and differ only in terms of the metaconceptual criteria examined the results for the two will be presented jointly.

EXPERIMENT 1

This experiment investigated three criteria for theory choice (see Samarapungavan, 1992 for a full description). Each of these metaconceptual criteria for theory evaluation and selection has been described extensively in the philosophy of science (Kuhn, 1977; Lakatos, 1978; Thagard, 1978). The criteria are described briefly below.

Empirical Consistency. According to this criterion a theory that is inconsistent with some empirical evidence that clearly bears upon that theory should be rejected in favor of a theory that is consistent with the empirical evidence.

Range of Explanation. According to this criterion all other things being equal, the theory that can account for more observations in a domain should be preferred over a counterpart that explains a more limited set of observations. This is true even though none of the phenomena circumscribed by the broader theory directly disconfirm the narrower theory.

Conceptual Coherence. According to this criterion, a theory that contains mutually contradictory propositions should be rejected in favor of an internally consistent theory.

Method

Subjects. The subjects were 150 elementary school children in Urbana, Illinois. There were three age groups: 1st graders (n = 50, mean age 6;9, range 6;0 to 7;5), 3rd graders (n = 50, mean age 8;5, range 7;8 to 9;3), and 5th graders (n = 50, mean age 10;11, range 10;0 to 12;1).

Task. Children's ability to use the above discussed criteria as a basis for selecting among alternate explanations was tested as follows: First the children were shown some set of observations. Then they were given two explanations that attempted to deal with those observations. Finally the children were asked which of the explanations they preferred and why. The two explanations presented for selection on each theory choice task were constructed so as to minimize differences in the plausibility of the conceptual content of competing theories for the children. Consequently the theories were "fake" theories and were not intended to reflect current scientific models.

The two theories were presented to the subjects as the theories of two other children, "Ann" and "Joe," who like to observe things around them and try to figure out why they happened.

Materials. Phenomena from two domains, those of astronomy and chemistry, were employed in the construction of the theory choice tasks. In the domain of astronomy, prior research (Vosniadou and Brewer, 1992) has shown that first and third grade elementary school children often explain phenomena such as the day/night cycle, seasons, and eclipses from a geocentric rather than heliocentric conceptual framework. By fifth grade most children have shifted to a heliocentric framework. This information was used to construct two separate sets of theory choice tasks for astronomy. In the first set to be referred to as the "geocentric" set each pair of theories was constructed so that their conceptual content would be compatible or neutral with regard to the conceptual frameworks of geocentric children. In the second set, to be referred to as the "heliocentric" set each pair of theories was constructed so as to be compatible or neutral with regard to the conceptual framework of heliocentric children.

Elementary school children were then pre-tested using an **astronomy questionnaire** (see Samarapungavan, 1992) to determine if they had heliocentric or geocentric conceptual frameworks for astronomy. Children were classified as geocentric if they indicated that the sun and moon moved relative to the earth and located the earth at the center of a picture of the solar system. They were classified as heliocentric if they indicated that the earth and mon moved around a stationary sun and located the sun at the center of a picture of the solar system. Each child was then presented with both the geocentric and the heliocentric set of theory choice tasks.

On the basis of the pragmatic model of scientific reasoning described earlier, it was predicted that the children should be able to select the better theory in all tasks where the content of both the theories was compatible with their conceptual framework. Therefore, geocentric children should be able to select the empirically consistent theory, the theory of broader range, and the conceptually coherent theory, on the geocentric theory choice tasks even if they failed to do so on the knowledge incompatible heliocentric theory choice tasks. Conversely, heliocentric children should be able to select the better theory on the heliocentric tasks even if they failed to do so on the geocentric tasks.

For purposes of generalizability, a third set of theory choice tasks centered around acid and base phenomena in chemistry was also presented to the children. As the children had no prior exposure to this class of phenomena the domain could be conceived of as knowledge neutral. Consequently, the children's prior knowledge was not directly tested in this domain.

Three sets of theory choice tasks were developed for each criterion (Geocentric Astronomy, Heliocentric Astronomy, and Chemistry). Thus, there were nine theory choice tasks and each child performed all the tasks. To illustrate the methodology used three of the tasks are described briefly below. For an extended description of these tasks see Samarapungavan (1992). On each task T1 always refers to the poorer theory on the metaconceptual criterion concerned and T2 to the better theory.

1. Heliocentric Astronomy Task - Empirical Consistency Criterion: The observations were four pictures of the night sky showing the phases of the moon (full moon, half moon, crescent moon, and no moon). As in the preceding example, both theories were initially equally successful at explaining the observations presented. However, new information provided by "Mr. Astronaut" after the theories had been presented supported the premises of T2 and undermined those of T1. On this task T1 and T2 were explained with the aid of physical models.

T1: The moon looks very different because every now and then, dark, black, rain clouds begin to form over the moon and block it. These clouds grow bigger and bigger until they cover all of the moon. Then the clouds dissolve in rain and we can see the whole moon again.

T2: The moon has a light and a dark side. The moon spins around above us and as it spins we see the half that is light and it looks round to us, As it turns, the dark side moved towards and we see only a small part of the light side, so the moon looks like a banana. As the moon continues to spin, only the dark side of the moon faces us and we cannot see it from earth.

Empirical test: Let us listen to Mr. Astronaut, who was up on the moon in his spaceship for two months. He will tell us what he saw there.

Mr. Astronaut: The moon is very different from earth. It was very dry on the moon indeed. We could find no water anywhere on the moon and were glad we had taken water with us from earth. The sky above the moon was so clear that we could see the stars every day. Our rocket had landed on smooth, white rock. Later as we walked over the moon we came across a large patch of black, dusty ground.

2. <u>Chemistry Task - Range of Explanation</u>: The observations consisted of 5 jars containing liquid acids or bases mounted on boxes that were labelled either "hot" or "cold." Two jars contained blue liquid and were mounted on "cold" boxes, two jars contained red liquid and were mounted on "hot" boxes, and one jar contained a colorless liquid and was mounted on a "cold" box. A pH indicator stick was dipped into each of the liquids and its color change was noted. The stick turned blue in the two jars with blue liquid and also in the jar with the colorless liquid. The stick turned red in the two jars with the red liquid.

The narrower theory (T1) proposed that the blue and red liquids were dyes which coated the stick but could not explain the stick's change of color in the colorless liquid. The broader theory (T2) proposed that the function of the stick was to indicate the temperature of substances and that it turned red in hot substances and blue in cold ones. This theory could thus also account for the observation that the stick turned blue in colorless liquid because the jar was mounted on a "cold" box.

T1: I think the stuff in the jars is paint. The stick is coated with the color of the liquid. So when you put the stick in the blue paint it turns blue and when you put the stick in the red paint it turns red.

T2: I think that the stick changes color to show if a thing is hot or cold. So when the liquid in the jar is heated by the hot box, the stick turns red. When the jar is on a cold box, the stuff in it is cold so the stick turns blue. That is why the stick turned blue in the jar with the liquid that had no color.

3. <u>Geocentric Astronomy Task - Conceptual Coherence</u>: The children watched while a rod and a wooden square were released from a small distance above the ground and allowed to fall. Then a small helium balloon was released and allowed to rise. The internally inconsistent theory (T1) proposed that the force of the earth pulled objects towards it causing the rod and the square to fall down, but it then explained the rising of the balloon by saying that the force of the earth propelled the balloon upwards away from it. The internally consistent theory (T2) used an analogy between air and water suggesting that the air could support things lighter but not heavier than itself. The rod and square which were heavy therefore fell downwards while the balloon which was light floated in the air.

T1: The rod and the wooden square fell when we let go of them because the earth has a force of its own. This force pulls all things down, back to the ground, when they are in the air. The balloon is going up because the force of the earth is pushing it away from the ground toward the sky.

T2: The air is like water. Some things are heavy and fall to the bottom but other things are light and float on top. If you let go of things that are big and heavy like a rock or wood, they will sink or go down in the air because the air is too light to hold them up. The balloon is light, so it can float on the air instead of falling to the ground.

Controls. The order of presentation of T1 and T2 per task was randomized for the nine tasks. The assignment of the better theory (T2) to "Ann" or "Joe" was randomized.

Scoring. 1. **Theory choices :** Whenever a child selected the appropriate theory on the basis of the metaconceptual criterion being tested the response was coded "Right." The selection of the inappropriate theory was coded "Wrong." A failure to choose between the theories or a rejection of both theories was coded "NA" or not applicable. 2. **Justifications of choice**: Children's justifications for their choice of theory were assigned to one of three categories. The first category, designated "**criterion-based**" represented a justification with explicit reference to the differences between theories on the metaconceptual criterion in question. For example, on the chemistry, range of explanation task (described above) one child selected the broader theory and said that T2 the broader theory was right because it "also shows why the stick is blue in that box - (points to clear jar)" while T1 did not. The second category of justification, designated "**content-based**" was based on the presumed "truth" or "falsehood" of the conceptual content of the theories. For example, with regard to the same task as in the earlier example one child selected the narrower theory and said that he knew it was "right about the paint because I studied about food color ... I don't think hot and cold is right." Children were assigned to this category based on their assertion that some aspect of the theories was either factually correct or incorrect. Whether the child's assertion itself was scientifically correct was not considered in the classification. The final category was that of "**no justification**." Children in this category did not justify their choice of theory.

EXPERIMENT 2

This experiment examined elementary school children's use of the non-ad hocness criterion. According to several philosophers of science (Lakatos, 1978; Popper, 1959; Thagard, 1978), theories that account for a set of phenomena in a simple and non- ad hoc way should be preferred over those that treat some observations in the set in an ad hoc way. Popper refers to ad hocness as the inclusion in a theory of auxiliary assumptions which are not independently testable and whose sole purpose is to save the theory from the specific empirical anomalies confronting it at that point in time.

As mentioned above, the general methodology and design for Experiment 2 were the same as for Experiment 1 and will not be described again.

Subjects. The subjects were 90 elementary school students in Urbana, Illinois. There were 30 first graders (mean age 6;6, range 6;3 to 7;1), 30 third graders (mean age 8;3, range 7;2 to 9;4), and 30 fifth graders (mean age 10;7, range 10;1 to 11;10).

Materials. Each child was given three theory choice tasks, one from each set of materials (i.e., geocentric astronomy, heliocentric astronomy, and chemistry). The tasks used were modified versions of those used in Experiment 1 to test the range of explanation criterion. The modification comprised of an ad hoc auxiliary assumption which was added to T1 the narrower theory on the range of explanation tasks to help it cope with data that it could not explain on the basis of its initial assumptions.

For example, the chemistry task used the same set observations employed in the chemistry - range of explanation task in Experiment 1 (see above). The task was set up so that initially the two explanations for these phenomena (see T1 and T2, chemistry - range of explanation) corresponded to a narrower theory which failed to explain why the stick changed color in the colorless liquid, and a broader theory that could account for all the phenomena in terms of a single mechanism. T1, the narrower theory then attempted to deal with the observation of the stick changing color in the colorless liquid by introducing the ad hoc assumption that some of the sticks got spoiled in storage and developed spots of color at random.

RESULTS

Table 1 shows the frequency of responses (pooled across grade) that children gave on the theory choice tasks. Overall, the frequency of correct choices (1267/1520) was very high. Only 275 incorrect choices were made and there 78 "NA" responses. A series of chi square analyses were performed on the raw data displayed in Table 1 to determine if the selection of appropriate theories on each theory choice task was significantly greater than chance. In these analyses the "NA" responses were excluded from consideration. Table 2 shows the results of the chi square analyses. On each of the nine tasks, the children did show a significant ($p<.01$) preference for the theory designed to be better in terms of the criteria being tested.

Table 1. Children's Theory Choices in Experiments 1 and 2

Criterion	Task	Theory Choice: Right	NA	Wrong	n
Empirical	Geocentric	130	14	6	150
Consistency	Heliocentric	142	1	7	150
	Chemistry	137	0	13	150
Range of	Geocentric	96	39	15	150
Explanation	Heliocentric	91	7	52	150
	Chemistry	140	1	9	150
Conceptual	Geocentric	120	1	29	150
Coherence	Heliocentric	120	4	26	150
	Chemistry	120	5	25	150
Non-	Geocentric	57	4	29	90
Ad Hocness	Heliocentric	51	2	37	90
	Chemistry	63	0	27	90
Total		1267	78	275	1620

Note. Data for subjects pooled across grades.

An analysis of justifications showed that most children gave content-based justifications when they chose the inappropriate theory with regard to the metaconceptual criterion being tested. There were however, interesting differences in the justifications provided for correct theory choices across the three criteria examined. In the empirical consistency condition, 96% of the justifications for correct theory choices were criterion-based while only 4% were content-based. Some of the most advanced criterion-based justifications observed in this study were in the empirical consistency condition. For example, one boy who chose the empirically consistent theory on the chemistry task said, "Joe is right because we did an experiment and we tested what he said and what Ann said. When we tested it, we saw that Joe came out right because it happened like he said it would. Ann came out wrong." In the range of explanation condition, 86% of all correct choices were justified on the basis of the metaconceptual criterion. Only 10% of the judgments were justified on the basis of conceptual content and 4% were not justified. In contrast, despite the preference for the conceptually coherent theory demonstrated by the children in their theory choices, many children (76%) did not provide any justification for their choice of the coherent theory. A small number of those who chose the conceptually coherent theory (22%) gave content-based justifications for their choices. It appears that although internally consistent explanations do not "make sense" to children and are therefore rejected (several children claimed that the inconsistent theory was "silly" or "weird") in favor of a coherent alternative, the conceptual coherence criterion is hard for children to articulate.

Very few of the children children (10%) who chose the non-ad hoc theory gave a criterion-based justification for their choice. All of the children who did give criterion-based justification for preferring the non-ad hoc theory were fifth graders. An example of a criterion-based justification is the response of a 5th grade child on the chemistry task. Having selected the non ad hoc theory (see materials section above) he said, "Joe is wrong 'cause you just can't say that it's (the stick) spoiled if it comes out colored when it isn't supposed to. So I think Ann is probably right about this." However, 57% of those who chose the non-ad hoc theory gave content-based justifications for their choice while 33% did not justify their choice of the non-ad hoc theory.

Table 2. <u>Chi Square Test for Probability of Correct Responses per Task</u>.

Criterion	Task	p_j	n	Chi Sq
Empirical	Geocentric	.96	136	126.36**
Consistency	Heliocentric	.95	149	109.97**
	Chemistry	.91	150	100.86**
Range of	Geocentric	.87	111	93.44**
Explanation	Heliocentric	.64	143	11.21**
	Chemistry	.94	149	115.39**
Conceptual	Geocentric	.81	149	95.36**
Coherence	Heliocentric	.82	146	39.48**
	Chemistry	.83	145	63.16**
Non-	Geocentric	.66	86	6.74**
Ad hocness	Heliocentric	.58	88	7.92**
	Chemistry	,70	90	14.40**

<u>Note</u>. Data for subjects pooled across grades.
**$p < .01$.

<u>The effects of knowledge and grade on the use of criteria</u>

Table 3 displays the percentage of correct choices by criterion and by grade for the knowledge consistent and knowledge inconsistent conditions in astronomy and for the chemistry tasks. The derivation of the performance measure for the knowledge-consistent and inconsistent conditions in astronomy is described below.

Based on the astronomy pretest, each child was categorized as having either a geocentric or heliocentric knowledge framework. This information was combined with information about each child's performance on the geocentric and heliocentric astronomy tasks. In the knowledge-consistent condition (Astronomy-C) the responses of geocentric children on geocentric tasks and those of heliocentric children on heliocentric tasks were pooled together. In the knowledge-inconsistent condition (Astronomy-I) the responses of geocentric children on heliocentric tasks and those of heliocentric children on geocentric tasks were pooled together. The results of the astronomy pretest were as follows. In grade 1, there were 12 heliocentric children and 38 geocentric children. In grade 3, there were 32 heliocentric children and 18 geocentric ones. In grade 5, all 50 children were heliocentric.

For each criterion, repeated-measures categorical analyses of variance were performed to determine the effects of knowledge and grade on theory choice in astronomy and the effects of grade on theory choice in chemistry. The results of these analyses will be discussed separately for each criterion.

Table 3. Percentage of correct responses by criterion and by grade.

		Condition		
Criterion	Grade	Astronomy-C	Astronomy-I	Chemistry
Empirical	1	98	92	88
Consistency	3	98	96	94
	5	96	94	94
Range of	1	78	20	90
Explanation	3	94	55	94
	5	98	88	98
Conceptual	1	84	60	71
Coherence	3	88	76	83
	5	98	82	94
Non-	1	70	35	43
Ad Hocness	3	63	50	77
	5	85	70	90

Empirical consistency. None of the variables had any effect on the use of the empirical consistency criterion as a basis for theory choice. This was because children performed very well at all grade levels on all three tasks. Even in grade 1 over 90% of the children chose the empirical consistent theory on the astronomy tasks and 88% did so on the chemistry task. Over 90% of the third and fifth graders also chose the empirically consistent theory on each task.

Range of explanation. There was no significant difference in performance on the chemistry tasks between grades. Even at grade 1, 90% of the children chose the theory of broader range. However, on the astronomy tasks both knowledge ($X^2(1, N = 150) = 18{,}56$, $P<.01$) and grade ($X^2(1, N = 150) = 9.61$, $p<.01$) had a significant effect on performance. At grade 1, 78% of the children chose the broader theory in the knowledge consistent condition but only 20% did so in the knowledge inconsistent condition. Indeed, in the latter condition a significant number of the first graders chose the narrower theory. Since this was the only case in which children's choices deviated significantly in the wrong direction, the children's protocols were re-examined for a possible explanation of these results. It turned out that all of the first grade children who had chosen the narrower theory in the knowledge-inconsistent condition were geocentric children who chose the narrower heliocentric theory. The justifications provided by these children showed that they found the premises of the narrower heliocentric theory (see Task 4 above) to be more plausible. The narrower theory explained the seasons in terms of cyclic changes in the amount of energy produced by the sun while the broader theory explained the seasons in terms of the revolution of the earth and the moon around the sun. Since geocentric children did not believe that the earth revolved around the sun, they found the narrower heliocentric theory to be more plausible.

At 3rd grade, 94% of the children chose the broader theory in the knowledge consistent condition although they performed at chance level in the knowledge inconsistent condition. At 5th grade 98% of the children chose the broader theory in the knowledge consistent condition. The effect of grade was not significant in the chemistry domain with as many as 90% of the 1st graders choosing the broader theory.

Conceptual coherence. The effects of grade and knowledge were not significant for the astronomy tasks. With one exception (1st graders in the knowledge-inconsistent astronomy condition) at each grade, over 70% of the children rejected the inconsistent theory in favor of the conceptually coherent theory. In the chemistry condition the effect of grade was significant (X^2 (2, N = 150) = 6.25, $p<.05$). As can be seen from Table 3, performance on the chemistry task improved with grade.

Non- ad hocness. The effects of knowledge (X^2 (N = 90) = 6.44, $p < .01$) and grade X^2 (N = 90) = 27.83, $p < .01$) on performance in the astronomy tasks were significant. Generally, third and fifth graders performed better than first graders. Except in the knowledge-consistent astronomy condition, both first and third graders did not systematically prefer non-ad hoc theories to ad hoc ones (see Table 3). However, fifth graders did prefer non-ad hoc theories in both astronomy conditions. The effect of grade was also significant in the chemistry condition (X^2 (N = 90) = 16.50, $p < .01$).

General Discussion

Overall, the results of the two experiments show that children know a good deal about several metascientific aspects of theory evaluation. Experiment 1 shows that on theory choice tasks where both theories are equally plausible, even children in the first grade can apply three different metaconceptual criteria for evaluating how competing theories. Further, at least in the case of the range and empirical consistency criteria, children could also articulate the metaconceptual principle involved as the basis for their theory preference. A second finding however, is that younger children (first graders) are significantly more likely to apply metaconceptual criteria, such as the range of explanation criterion, to theory selection in cases where the conceptual content of the rival theories was either compatible or neutral with regard to their domain specific beliefs than in cases where the content of both theories was incompatible with their beliefs. Further, if a plausible theory is pitted against an implausible one their choice tends to be based on the conceptual content rather than on any metaconceptual differences between the two theories. Older children (third graders and fifth graders) are less influenced by the compatibility of the theories with their prior knowledge.

The only criterion which first and third graders could not systematically apply to all the theory-choice tasks was the non-ad hocness criterion. This criterion is a complex one. Pragmatic philosophers of science like Laudan (1977) have noted that modifications to a theory that appear ad hoc to start with have sometimes led to important novel insights. Perhaps judgments of ad hocness are also dependent on children's developing knowledge of domains of explanation and the construction of domain boundaries, since such judgments require decisions about what kinds of explanatory mechanisms should be applicable to different classes of phenomena.

Such findings point to the intricate relationship between domain knowledge and metacognitive knowledge. The findings are consistent with the pragmatic model of scientific reasoning elaborated in the philosophy of science literature showing that once some conceptual content has been assigned a great degree of certainty, it is becomes difficult (though not impossible) to eliminate or modify that conceptual content on the basis of general metaconceptual principles. I should note that these experiments point to a somewhat different role for prior knowledge in influencing problem solving and reasoning performance from that which is the traditional focus of the cognitive developmental literature. It is widely acknowledged that the difficulty level of a cognitive task depends on the degree of prior knowledge a subject has in the task domain and that people perform poorly with tasks that require domain specific knowledge that subjects lack. For example, one would find it difficult to evaluate the empirical success of Ptolemaic versus Copernican models of the solar system if one was ignorant of the data or observations that each model could explain. However, these experiments show that domain specific knowledge also affects our judgment about the plausibility of new ideas or hypotheses even when these hypotheses are comprehensible to us.

I suggest that certain kinds of metascientific knowledge which are manifest in a general ability to monitor the internal consistency of mental representations and the fit of these representations to the external world appear early in development and may even be built into the cognitive machinery. However, it is not my intention to argue that children have nothing to learn about metascience. Indeed, this basic metacognitive knowledge can be, and most likely is, refined by learning and instruction. For example, the professional scientist learns the culturally shared criteria for what kinds of evidence are adequate to bring about consensus or to persuade one's peers. Some of these criteria may be considered as rather general metaconceptual criteria such as the reliability or replicability of the evidence, while others maybe more domain specific such as knowing which variables one ought to control for, in trying to determine causal relations, having access to data reduction and data analyses techniques and so forth. Many of the more domain specific criteria for evaluating evidence have changed historically for professional scientists over time. Of course there is a sense in which the professional scientist is much likelier to succeed in producing "better" theories on such metaconceptual criteria as empirical consistency and parsimony than the child. But a good part of this superior performance can be attributed to the scientist's superior knowledge of the domain, its methodological conventions, and last but not least, to the scientist's liberation from the limitations of short term-memory through good record keeping.

As a final note, the results of this research are in my view relevant to the broad debate about the nature of conceptual change and how conceptual change is to be directed in science education. The research suggests that children can evaluate explanations on metaconceptual criteria that play an important role in scientific reasoning when these criteria are made salient to them. However, current science instruction rarely contrasts alternative explanations or theories along such metaconceptual dimensions in presenting science concepts to children. Instruction which supports an active use of such metaconceptual evaluative criteria in science learning would serve two goals. Firstly, it would foster an explicit awareness and systematic application of metaconceptual knowledge in doing science. Secondly, it would help reduce "misconceptions" in students by encouraging them to contrast their own ideas with expert scientific concepts on metaconceptual criteria of the kind examined here.

REFERENCES

Brewer, W. F., & Samarapungavan, A. (1991). Child theories versus scientific theories: Differences in reasoning or differences in knowledge. In R.R. Hoffman & D.S. Palermo (Eds.), *Cognition and the symbolic processes: Vol. 3. Applied and ecological perspectives* (pp. 16-32). Hillsdale, NJ: Earlbaum.

Brown, A. L. (1990). Domain-specific principles affect learning and transfer in young children. *Cognitive Science, 15,* 107-134.

Carey, S. (1985). Are children fundamentally different kinds of thinkers and learners from adults? In S. Chipman, J. Segal, & R. Glaser (Eds.), *Thinking and learning skills: Research and open questions* (Vol. 2, pp. 485-517).

diSessa, A. (1988). Knowledge in pieces. In G. Forman, & P. B. Pufall (Eds.), *Constructivism in a computer age* (pp. 49 - 70).

Driver, R. & Easley, J. (1978). Pupils and paradigms: A review of the literature related to concept development in adolescent science students. *Studies in Science Education, 5,* 61-84.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23,* 183-208.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*, 674-689.

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

Kuhn, T.S. (1977). *The Essential Tension*. Chicago: University of Chicago Press.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.

Laudan, L. (1977). *Progress and its problems*. Berkeley, CA: University of California Press.

Laudan, L. et al. (1986). Scientific change: Philosophical models and historical research. *Synthese, 69,* 141-223.

McCloskey, M., & Kargon, R. (1988). The meaning and use of historical models in the study of intuitive physics. In S. Strauss (Ed.), *Ontogeny, phylogeny, and historical development* (pp.49-67). Norwood, NJ: Ablex.

Samarapungavan, A. (1992). Children's' judgments in theory choice tasks: Scientific rationality in childhood. *Cognition*, *45,* 1-32.

Solomon, J. (1983). Thinking in two worlds of knowledge. In H. Helm & J. D. Novak (Eds.), *Proceedings of the International Seminar: Misconceptions in Science and Mathematics* (pp.127-132). Ithaca, NY: Cornell University.

Thagard, P. (1991). The conceptual structure of the chemical revolution. *Philosophy of Science, 57,* 183-209.

Vosniadou, S. (1989). Analogical reasoning and knowledge acquisition. In S. Vosniadou &A. Ortony, (Eds.), *Similarity and analogical reasoning* (pp. 413-437). Cambridge: Cambridge University Press.

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

Wiser, M. (1988). The differentiation of heat and temperature: History of science and novice-expert shift. In S. Strauss (Ed.), *Ontogeny, phylogeny, and historical development* (pp.28-48). Norwood, NJ: Ablex.