Generating Initial Models for Reasoning

James A. Dixon

College of William & Mary

and

Frank Tuccillo

Columbia University

Much of the knowledge that children and adults have about the world resides in intuitive models. Previous work shows that intuitive models allow for computation of specific outcomes given information about the system, but little is known about how such models are acquired. The current study tested three hypotheses about how children and adults construct intuitive models when they encounter a new property: (1) intuitive models are constructed by transferring principles from familiar properties; (2) with development, children shift from applying a default model to constructing specialized models; and (3) younger children’s model construction is constrained by the domain, but becomes increasingly domain independent with development. Participants from three age groups (10, 13, and 19 years) made a series of judgments about two familiar properties and one novel property. Causal models showed that all age groups transferred principles from the familiar properties to the novel property. None of the age groups used a default model. There was developmental change in the effect of domain; younger, but not older, children’s models were affected by domain. These findings suggest that the transfer process is developmentally invariant but that constraints on the process (i.e., domain dependence and understanding of base models) change with development. © 2001 by Academic Press

Key Words: reasoning; knowledge acquisition; analogical transfer; cognitive development.

Children develop an impressive understanding of many of the functional relationships that exist in the world. Even fairly young infants can identify that particular events are causally related to one another. For example, Oakes and...
Cohen (1994) showed that 6-month-old infants begin to perceive launching events (i.e., the movement that results from one ball striking another) as causal. A large body of research with older children shows that they not only identify causal relationships, they also construct models that predict how multiple factors affect the variable of interest. For example, Wilkening (1981) showed that children develop intuitive models of how the distance an object travels is affected by the velocity of the object and the time it spends in motion. Similar models are also found in children’s social reasoning. For example, Surber (1980) showed that children as young as 6 years of age have fairly good intuitive models of the effects of ability and effort on performance. Surber found that children’s models of performance specify not only that both ability and effort are important to performance, but also how ability and effort combine to affect performance.

We refer to these models as intuitive or qualitative to distinguish them from the formal, usually mathematical, models scientists construct. In general, intuitive models specify how changes in particular properties within a system affect the property of interest. Properties may be defined as functionally important, and often perceptually salient, variables within a system. Systems are sets of interrelated properties. For example, suppose a child is making lemonade by mixing lemon juice and water. The quantity of water and the quantity and concentration of the lemon juice are examples of properties within the lemonade-making system.

Children and adults have sophisticated intuitive models of many different systems, spanning a broad variety of content domains. For example, people appear to have intuitive models of temperature mixture (e.g., Ahl, Moore, & Dixon, 1992), acid mixture (e.g., Reed & Evans, 1987), expected value (Schlotmann & Anderson, 1994), balance scales (e.g., Siegler, 1981; Surber & Gzesh, 1984), moral judgment (e.g., Anderson, 1990; Surber, 1977), health (e.g., Hermand, Mulett, & Coutelle, 1995), and motion (e.g., Kaiser, McCloskey, & Proffitt, 1986). Indeed, much of the knowledge that people have about their physical and social worlds seems to reside in the form of such intuitive models, although other types of knowledge representations are also important in reasoning such as category structure (see, for example, Huang, 1999). It is worth noting that intuitive models are sometimes systematically inaccurate in both children and adults (see, for example, Proffitt & Gliden, 1989; or McCloskey & Kohl, 1983).

How children come to acquire their complex intuitive models is not well understood. While many studies have documented children’s use of intuitive models and changes in those models with development, little is known about how children construct these models. The purpose of this article is to provide some preliminary evidence about processes children use to construct intuitive models when they encounter new properties. First, we briefly summarize some fairly recent evidence concerning the nature of intuitive models. Next, we discuss how children might construct intuitive models and propose three hypotheses about the model construction process.
The Nature of Intuitive Models

Past research has demonstrated that intuitive models are not sets of simple declarative propositions. Although children and adults can sometimes summarize relationships within their intuitive models as propositions (e.g., “the longer the lever, the greater the force”), the fact that they can use their intuitive models to integrate variables so as to predict a particular outcome strongly suggests that intuitive models are computational rather than propositional. While “the longer the lever, the greater the force” is true for lever systems, it is not sufficient to provide an estimate of how much force is generated by a system under a particular set of conditions (cf. Anderson, 1987). Of course, the assertion that intuitive models are computational does not imply that children cannot access other representations that are noncomputational.

Compelling empirical evidence for the computational nature of intuitive models emerges when children make a series of predictions, usually completing a factorial design. The resulting pattern of responses resembles the pattern one would obtain by combining the variables using an algebraic rule. For example, Moore, Dixon, and Haines (1991) presented participants with two containers of water. Each container had a specified quantity of water at a particular temperature. The participants were asked to estimate what the temperature of all the water together would be for each of 25 randomly ordered combinations. Children’s patterns of estimates were systematically related to the variables being combined, with some children producing patterns of estimates that were consistent with weighted averaging, the normative algebraic rule. Other children produced estimates that were consistent with less sophisticated algebraic rules, such as unweighted averaging. The point here is not that children have the appropriate mathematical algorithm—indeed it is quite clear that the children were not using mathematics in this task (see Dixon & Moore, 1996). Rather their intuitive models are “computable”; that is, they function like an algebraic algorithm.

Recent evidence suggests that intuitive models are “computable” because they are composed of principles. Each principle specifies a particular relationship in the

FIG. 1. The top panel shows an example of a mixture task in which two containers each hold some quantity of a substance. To illustrate the averaging model, suppose the substance is water and that property of interest is temperature. Each container has water of a different temperature. The intensity of the temperature is represented by the shading of the water. The contents of the container on the right, the “Added Container,” is added to the container on the left, the “Initial Container.” The task is to judge the resulting intensity of the temperature in the left container; “What is the temperature of all the water together?” To illustrate the additive model, suppose the containers hold different types of metal; aluminum on the left, lead on the right. The property of interest is weight. Aluminum is lighter than lead. This difference in the intensity of the weight is represented by the shading of the metal. The task is to judge the resulting intensity of the weight; “What is the weight of all the metal together?” The remaining panels show fuzzy set representations of the principles of the averaging and additive models. Intensity (e.g., of temperature or weight) is on the horizontal axis. Each principle specifies a fuzzy set of potential answers (abscissa). Membership in the fuzzy set is given on the ordinate. $I_i$ refers to the initial intensity; $I_a$ refers to the added intensity. $Q_i$ refers to the initial quantity; $Q_a$ refers to the added quantity. $R1$ refers to the answer to a previous problem. The individual principles are described in the text.
model. For example, children have a sophisticated model of the temperature mixture task described above and depicted at the top of Fig. 1. The container on the left, which we refer to as the initial container, holds a particular quantity of water ($Q_i$) at a particular temperature ($I_i$). (Because we discuss these principles with reference to other properties later in the article, we use "$I" to stand for the more general term...
“intensity”; here referring to the intensity of the temperature. The shading indicates the intensity of the temperature.) The container on the right, the added container, also holds a particular quantity of water ($Q_a$) at a particular temperature ($I_a$). If the contents of the added container is poured into the initial container, what is the temperature of all the water together? Children represent this problem with a group of principles, including “range,” “monotonicity,” and “quantity.”

The range principle for temperature mixture specifies that the final temperature must be between the two combined temperatures, $I_i$ and $I_a$. More precisely, the range principle specifies a fuzzy set of potential answers given the two temperatures being combined (Zadeh, 1965). The panel labeled “Range: Averaging Model” in Fig. 1 shows an example of the fuzzy set for this principle. Potential answers are on the horizontal axis. The degree of membership in the fuzzy set for each value is shown by the height of the curve. The members of the fuzzy set are all between the two temperatures, $I_i$ (the initial temperature) and $I_a$ (the added temperature), but potential answers nearer the center of the region have the strongest membership. Note that the principle represents semantic knowledge about the resulting temperature being between the two combined temperatures rather than logical knowledge (Dixon & Moore, 1997). On logical grounds, all values greater than $I_i$ and less than $I_a$ would be perfect members of the set “between.” However, considered from a more semantic perspective, values toward the center of the region are more characteristic of being “between” and therefore better members of the set. [See Wallsten, Budescu, Rapoport, Zwick, & Forsyth (1986) for similar findings with probability concepts and Zadeh (1983) and Zimmer (1988) for similar conclusions regarding linguistic quantifiers.]

The monotonicity principle for temperature mixture specifies that the greater the temperature of the added water, the greater final temperature. A fuzzy set representation of this principle is shown in the panel labeled “Monotonicity: Additive and Averaging Model” in Fig. 1. Again, the potential answers are along the horizontal axis; the membership of each value in the fuzzy set is shown by the height of the curve. Assume that the child has already made a judgment about a prior problem in which the added container held water at some temperature, $I_{a1}$. The child’s response for that problem is represented on the horizontal axis as $R_1$. A second problem is then presented in which the added container holds water at a higher temperature, $I_{a2}$. The monotonicity principle specifies that only values that are greater than $R_1$ are members of the fuzzy set of potential answers to this second problem.

The quantity principle for temperature mixture specifies the interaction between quantity and temperature; the greater the quantity of the water, the greater the effect of its temperature. The panel labeled “Quantity: Averaging Model” shows a fuzzy set representation of this principle. Assume that the quantity of the added container ($Q_a$) is greater than the quantity of the initial container ($Q_i$). Potential answers that are closer to the added temperature ($I_a$) are members of the fuzzy set of potential answers. The other fuzzy set representations in Fig. 1 are described later. [See Ragade & Gupta, (1977) and...
Zadeh (1965) for discussions of fuzzy set theory. See Massaro (1994) or Oden (1984, 1988) for examples of fuzzy set theory applied to modeling cognitive and perceptual processes.

Taken together, these principles comprise the averaging model. Judgments about individual problems are computed by combining the separate principles. The principle representation hypothesis is supported by a variety of converging lines of evidence. First, Reed and Evans (1987) demonstrated that directly stating principles of acid mixture, such as range, to participants was a more effective instructional technique than either presentation of examples or graphic displays of the relationships. Specifying which set of principles was relevant greatly facilitated problem solving performance.

Second, Moore et al. (1991) and Ahl et al. (1992) showed that children’s performance on the temperature mixture task is explained by the acquisition of five principles, including the ones described above. Children acquire the principles in one of two developmental sequences.

Third, children appear to access the principles during problem solving. For example, Dixon and Moore (1996) found that children who mentioned a principle while they were attempting to solve temperature mixture problems with mathematics had produced estimates consistent with that principle in a prior estimation task. In a subsequent study, Dixon and Moore (1997) found that when children and adults were asked to verify answers to problems, they rejected answers that violated a principle more quickly and accurately than answers that did not violate a principle but were equally distant from the correct answers. In addition, participants who received instruction about a principle were better able to verify answers to problems. Interestingly, the effect of instruction on verifying answers was mediated by measures of understanding the principles. That is, it appears that instruction affected the participants’ representation of the principles which, in turn, affected their ability to verify answers. Taken together, these lines of evidence support the conclusion that the intuitive model is composed of principles. While the principle representation hypothesis describes the models children eventually acquire, it does not explain the acquisition process. Indeed, very little is known about how children construct their intuitive models.

Constructing Intuitive Models

Children are believed to construct models for some properties very quickly and easily because their perceptual and cognitive systems have been specialized by evolution to detect and interpret the relevant relationships. Researchers have identified these properties as “privileged.” For example, this argument has been advanced for numerosity as well as other properties in the number domain (e.g., Gallistel & Gelman, 1992; Wynn, 1992). Children are thought to be predisposed to acquire the appropriate models for these properties with only limited experience.

However, most properties are not believed to have privileged status. This makes considerable sense given that people develop intuitive models for a vast array of properties that are poor candidates for specialization from evolution (i.e.,
properties that are culturally specific, recently discovered, or of modest survival value). For example, children develop models of the Earth’s shape (Vosniadou & Brewer, 1992), a recently discovered property (at least on an evolutionary scale) that is accepted only in some cultures and of little survival value.

Models for nonprivileged properties are thought to be acquired using a set of general-purpose mechanisms. Research on knowledge acquisition suggests that these mechanisms generate a candidate or initial model for the new property and subsequently evaluate whether that model is appropriate as evidence is collected (Khlar, Fay, & Dunbar, 1993; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995). For example, Halford, Brown, and Thompson (1986) found that children, ages 7–14 years, at first generated a simple initial model for the property “flotation” (i.e., will an object float in water). The majority of their models consisted of a single relationship: Lighter objects float, heavier objects sink. After evidence was presented about whether particular objects floated, nearly all the older (11–14 years of age) and most of the younger (7–9 years of age) children modified their initial models.

We were particularly interested in how children and adults generate an initial model because past research suggests that the initial model exerts a powerful influence on which model is eventually accepted. People tend to gather evidence only for models that they consider plausible rather than explore the entire range of possible models (Dunbar, 1993; Khlar et al., 1993; Kuhn et al., 1995). Similarly, the evidence they gather is used to confirm rather than disconfirm models (Klayman & Ha, 1987). In addition, children sometimes retain a model despite having evidence that disconfirms it (Schauble, 1996). Because children and adults use their initial intuitive model to guide their search for evidence, as well as their interpretation of that evidence, the initial model will play a critical role in constructing the final intuitive model.

Note that generating an initial model for a new property is different from generating a new type of model (e.g., the averaging model). The initial model for a new property is generated before evidence about the property is collected and, therefore, must be based primarily on past experience with other properties. In contrast, generating a new type of model, a model that contains completely new relationships, seems necessarily based on the evaluation and interpretation of evidence about those new relationships. For example, suppose a child encounters the property “temperature” for the first time. Unless the child is familiar with other properties that follow this type of model (i.e., the averaging model), his or her initial model for temperature is very unlikely to contain the appropriate relationships. However, given sufficient experience with temperature the child may discover the appropriate relationships, using processes like gist extraction from fuzzy-trace theory (Reyna & Brainerd, 1995). Eventually the initial model would be modified into a new type of model that includes the new relationships.

In the current study, we tested three hypotheses about how children and adults construct their initial models for new properties. The first hypothesis stems from the evidence discussed above regarding the nature of intuitive models. We hypothesized that when children and adults construct their initial models they
are actually transferring principles from previously understood properties. According to this hypothesis, the initial model for a new property is assembled from individual principles. These individual principles are taken from the intuitive models for familiar properties. One important developmental implication of this hypothesis is that the initial models people construct for new properties will be constrained by their models for familiar properties. That is, a principle can only be included in the initial model to the degree that the principle is present in the model for the familiar property. We predicted this process would be developmentally invariant across the age groups sampled (5th grade, 8th grade, and college).

The second hypothesis is that children use a particular model as a default whenever they encounter an unknown property. Just as experimental scientists first model phenomena with the simplest linear models, children might be expected to first employ a simple default model for new properties. This would allow them to conduct structurally similar tests across a wide variety of properties rather than simultaneously generate a tailored model and the tests necessary to evaluate that model. However, because people have a confirmatory bias toward the initial model, a potential cost of this approach is that the default model would often be erroneously accepted. We expected this strong constraint would be most pervasive for the youngest age group, but that older children would be more flexible in constructing their initial models.

The third hypothesis is that, rather than use a default model, children use the model that is most pervasive in the current domain as the initial model. Consistent with some previous uses of the term (e.g., Case & Okamoto, 1996; Springer & Keil, 1991), we define domain as a collection of systems that have similar objects in similar functional roles. [See Hirschfeld & Gelman (1994) for a discussion of definitions of domain.] For example, blocks and boxes would be in the same domain because they share similar physical characteristics (e.g., shape and rigidity) and functionality (e.g., stackability). As we explain in more detail later, different models are not equally salient across domains. According to the domain hypothesis, to the extent that children experience a model as predominant within a particular domain, new properties encountered in that domain will be assigned that model (see Mandler & McDonough, 1998). For example, suppose a child repeatedly uses the additive model while interacting with blocks. He or she observes through experimentation that the heights, lengths, and weights of blocks combine additively. When a new property in this domain is experienced, for example, buoyancy, the child would use an additive model as the initial model for that property. We expected younger participants to show strong effects of domain and older participants to show increasing domain independence.

We addressed these hypotheses in the context of reasoning about novel physical properties. Participants were asked to judge what happened to a novel physical property when two quantities of a substance were combined. Consider the following problem as an example. Two containers of water have different degrees of “hemry,” a novel property. Suppose one container holds very hemry water and a
second container holds slightly hemry water. What is the hemry of all the water together? Because hemry is a fictitious property, there is no objectively correct answer to this question. However, just as people approach experimental problems...
with an initial hypothesis in mind (Kuhn et al., 1995), we expected our participants to generate an initial model of how this novel property combines. While there are many potential models for combining properties, past research on people’s understanding of mixture tasks such as this one, strongly suggests that children and adults will use one of two models: an additive model or an averaging model.

The additive and averaging models. A remarkable number of properties combine according to the additive model. For example, height, weight, and force combine additively under certain conditions. When both the intensity of the property (e.g., weight) and the quantity of the substance (e.g., candy) are independently varied the result is a weighted addition model. The top panel of Fig. 2 shows the normative pattern of combined intensities for this model when the contents of one container are held constant and the contents of the other container vary in a 3 (Intensity) x 3 (Quantity) factorial design. The units on the vertical axis are from the judgment task described below, but may be considered arbitrary for the current discussion.

The averaging model is also very pervasive. Many properties such as temperature, sweetness, and acidity average when combined. When both the intensity of the property and the quantity of the substance are independently varied these properties follow a weighted averaging model. The bottom panel of Fig. 2 shows the normative pattern, again when the contents of one container is held constant and the contents of the other container varies factorially. We used a mixture task to investigate how people generate initial models because a limited number of combination models are likely and a strong methodology exists for distinguishing among models (e.g., Ahl et al., 1992; Anderson, 1976, 1981; Moore et al., 1991; Surber, 1984).

The research discussed above suggests that people represent the additive and averaging models with sets of principles. We focused on three of these principles: range, monotonicity, and quantity. The values of these principles for the averaging model were explained previously with reference to temperature mixture. Recall that for the averaging model, range specifies that the final intensity must be between the two combined intensities; the monotonicity principle specifies that the greater the intensity of the substance being added, the greater the intensity of the answer; and the quantity principle specifies the interaction between quantity and intensity. The relevant fuzzy set representations, shown in Fig. 1, were also described above.

As an example of the additive model, again consider the task depicted at the top of Fig. 1. However now suppose that, rather than water, the initial container holds some quantity of aluminum and the added container holds some quantity of lead. These metals differ in terms of their weight; that is, equal quantities of aluminum and lead weigh different amounts. Let the shading of the contents indicate the intensity of weight. If the contents of the added container are combined with the contents of the initial container, what is the weight of all the metal together?

For the additive model, range specifies that the final intensity (e.g., weight) must be greater than either of the combined intensities. A fuzzy set representation of this principle can be seen in the panel labeled “Range: Additive Model” in Fig. 1. Only answers with greater values than either intensity are in the fuzzy set specified by the range principle.
FIG. 3. An example of the judgment task. The top panel shows the presentation of two different types of drink that differ in the property sweetness. Each container has a corresponding “sweetness meter.” The position of the marker on the meter, and the description in the text, indicated the sweetness of the drink. The middle panel shows the presentation of a different quantity of the type of drink on the right. Participants adjusted the marker on the “sweetness meter” to indicate the sweetness of the new quantity. (The quantity change affects the property weight, but not sweetness). The bottom panel shows the presentation of the final question in three separate parts, labeled 1, 2, and 3, which were presented sequentially. The drink on the right has been transferred into the container on the left. The line around the left-side container indicates the original amount of drink. Participants adjusted the marker on the left-side meter to indicate the sweetness of all the drink together.
The monotonicity principle is the same for the additive and averaging models. The greater the intensity of the substance being added, the greater the intensity of the answer. The fuzzy set for this principle was explained previously with reference to temperature mixture.

For both models the quantity principles specifies the interaction between quantity and intensity, but the form of the interaction is different. As can be seen in the top panel of Fig. 2, in the additive model the effect of quantity increases as the added intensity increases. The panel labeled “Quantity: Additive Model” in Fig. 1 shows a fuzzy set representation for this principle. Assume that a child has already made a judgment for a previous problem which had an added quantity of $Q_a1$ and added weight of $I_a1$. The child’s answer is represented as $R1$. A second problem is then presented which has a greater added quantity, $Q_a2$, but the same added weight. Values that are larger than $R1$ are members of the fuzzy set of potential answers.

Overview of the experiment. Fifth-grade, 8th-grade, and college participants were asked to make judgments about three different properties. Two of these were familiar properties: weight and sweetness. The third was a novel, fictitious property, called “hemry.” Half the participants made judgments about each property in the drinkable-liquids domain, which we refer to as the “drink” domain. Participants assigned to the drink domain made judgments about the weight, sweetness, and hemry of combinations of drinks. The remaining participants made their judgments in the eatable-solids domain, which we refer to as the “candy” domain. These participants made judgments about the weight, sweetness, and hemry of combinations of candies. To explain the task, we use the familiar property “sweetness” and the domain “drink.”

For each property, participants were shown two containers. Each container held a substance, a drink in this example. The drink in each container was described as having a particular intensity or degree of the property (e.g., sweetness). For example, the top panel of Fig. 3 shows two containers of drink. The type of drink in the container on the left is fairly sweet, the type of drink in the container on the right is very sweet. As can be seen in the figure, this information about the property was given as text and was also indicated by the position of a marker on a “sweetness meter.” Next, participants were shown a different quantity of the type of drink on the right and asked to adjust the marker to indicate how sweet this much drink would be (middle panel of Fig. 3). (Properties

---

1 The reader may note that the quantity principle for the additive and averaging models have different terms on the horizontal axis. This highlights a difference in the quantity principle for these models. For the averaging model, the quantity principle specifies a region of possible answers relative to the intensities in the current problem; the answer will be closer to intensity of the container with the larger quantity. However, the quantity principle for the additive model only specifies a region of possible answers across problems. For example, knowing that the quantity of the initial container is greater than that of the added container does not specify a region of answers relative to the intensity of either container. But if the current problem is compared with another problem with different quantities, then the principle specifies the relative positions of the answers.
that follow an averaging model, such as sweetness, are not affected by changes in the quantity of the substance. Properties that follow an additive model, such as weight, are affected. Finally, participants were asked how sweet all the drink together would be if the two containers were combined (bottom panel of Fig. 3). Again, they responded by adjusting a marker on a sweetness meter. The sweetness and quantity of the drink on the right varied in a factorial design. There was always a medium amount of fairly sweet drink in the container on the left.

Participants were randomly assigned to one of two domains: drink or candy. We chose these two domains because drinkable liquids is a very familiar liquid domain and eatable solids is a very familiar solid domain. Each student made nine judgments for each of three properties (i.e., weight, sweetness, and hemry) in their assigned domain. Within each domain, the displays were identical across properties; only the name of the property changed. The only difference in the displays between domains was whether drink or candy was shown in the container.

This design allowed us to assess the participants’ models for the familiar properties, weight and sweetness, and for the novel property, hemry, in this task context. We expected participants in this age range to have considerable experience with these particular familiar properties. However, based on past research, we also expected to observe developmental and individual differences in employing the appropriate models for these properties.

Predictions. The first hypothesis, that children and adults construct their initial models by transferring principles from already familiar properties, predicts a systematic relationship between the principles for the familiar and novel properties. Principles for the familiar property should be correlated with principles for the novel property. Further, only principles that are structurally isomorphic (i.e., specify the same relationship within each property) should be uniquely related. That is, principles that specify the same relationship within the familiar and novel property should have significant semipartial correlations. We explain this prediction in more detail under Results with reference to the specific principles.

The second hypothesis, that children apply a default model when they encounter a new property, is based on past research which showed that young children first learn and overgeneralize the additive model (Strauss & Stavy, 1982). We expected the younger participants to use the additive model as a default and the older participants to use both models. Therefore, the younger children’s initial models for hemry should be very similar to their models for weight.

The third hypothesis is that younger children apply the model that is most pervasive in the current domain. Although the additive and averaging models are appropriate for properties in both the liquid (e.g., drink) and solid (e.g., candy) domains, the models are not used equally across these domains. In children’s and adults’ everyday experience, different liquids are formed by suspending different solutes in a base liquid or solvent. For drinkable liquids this base is almost always water. Because differences among liquids are usually based on differences in the concentration of a solute and concentrations average when combined, the aver-
aging model dominates in this domain. For example, properties such as clarity, acidity, sweetness, and viscosity, which are typically used to distinguish among liquids, all combine according to an averaging model.

Differences among solids are most often based on differences in extensive properties, such as weight, height, and length. Because the concentration of most properties in a solid are fixed by virtue of being “locked” in the solid, the concentrations of these properties rarely change in everyday experience with combinations. Therefore, the additive model, as opposed to the averaging model, is usually appropriate when solids are combined. For example, when a child combines two pieces of cake, he or she is very likely to consider their joint height or weight, properties that combine additively. However, the child is unlikely to consider properties that combine according to an averaging model, such as temperature.

The effect of domain on children’s initial models should be strongest early in conceptual development. Children first learn about properties in the context of specific domains. With development, understanding of a property is extended beyond its original domain. For example, the property “quantity” is presumed to be learned first in the context of discrete objects. Later, children’s understanding of this property is extended to continuous substances (cf. Fuson & Hall, 1983). As children’s understanding of a property becomes increasingly independent of the domain, the effect of domain on proposing an initial model should decrease. Therefore, we expect a developmental shift in the use of domain as a constraint. The predictions are summarized in Table 1.

METHOD

Participants

One hundred twenty-seven participants from three age groups, 5th grade (n = 50), 8th grade (n = 37), and college (n = 40), completed the experiment. 2 The mean age for each grade was as follows: 5th grade: 10 years, 6 months; 8th grade: 13 years, 9 months; and college: 19 years, 1 month. One additional 8th-grade student was eliminated from the experiment due to a computer hardware problem.

Design

Participants were asked to make nine judgments about each of three properties: weight, sweetness, and hemery. Each set of judgments was immediately preceded by three practice trials. Judgments about each property were blocked together. Participants were randomly assigned to one domain, either drink or candy. For participants assigned to the drink domain, each property described a quality of drinks, that is the weight, sweetness, or hemery of drinks. For participants assigned

2 We recruited more 5th-grade participants because past research has shown their responses to contain a larger error component. Therefore, in order to equate the probability of detecting the predicted effects (i.e., power) across age groups, a larger sample of 5th-graders was required (Maxwell & Delaney, 1990).
to the candy domain, each property described a quality of candy. Participants were also randomly assigned to one of two different order conditions. Approximately half of the participants first made judgments for weight, then for sweetness, and finally for hemry. The remaining participants first made judgments for sweetness, then for weight, and finally for hemry. Assignment to these orders counterbalanced the order of the familiar properties. The novel property hemry always appeared last so that participants had experience instantiating both models in the task context prior to making judgments about hemry.

Procedure

The experimenter read a short cover story about a group of imaginary scientists conducting research on either drink or candy. The scientists were preparing different types of drink or candy for commercial production and were attempting to find the kind that people would like best. Participants were asked to help the scientists figure out how different properties of either drink or candy worked. Participants were told that they would be asked to make judgments about two properties with which they were already familiar, sweetness and weight, and a third property that the scientists had just discovered. They were asked to make the best judgment they could for each question. The judgment task was identical across domains except that in one domain containers of drink were combined and in the other domain containers of candy were combined.

The task was presented on a Macintosh PowerPC computer using HyperCard 2.2. Participants were first told that a scientist wanted to combine two different types of a substance (e.g., two different types of drink). Figure 3 shows an example of the display. Each substance was shown in a container. The two substances

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principle Transfer</td>
<td>Principles for the familiar and novel property will be correlated but only isomorphic principles will share unique variance. The transfer process should be developmentally invariant.</td>
</tr>
<tr>
<td>Default Model</td>
<td>There will be a developmental shift from use of an additive model as the default initial model toward using both the additive and averaging models.</td>
</tr>
<tr>
<td>Domain</td>
<td>Because the averaging model predominates in liquid domains and the additive model predominates in solid domains, children’s initial models for properties of liquids will follow the averaging model and their initial models for properties of solids will follow the additive model. The effect of domain should decrease with development.</td>
</tr>
</tbody>
</table>

*Note. The predictions of the major hypotheses in the study are summarized above. Isomorphic principles specify analogous structural relationships.*
looked identical and the containers held equal amounts of the substances. When the containers appeared on the screen, a marker on a meter next to each container quickly moved to a position indicating the intensity of the property (i.e., weight, sweetness, or hemry) for each substance. The top end of each meter was marked as “Greatest ‘Property’” and the bottom end was marked as “Least ‘Property,’” where “Property” was weight, sweetness, or hemry. The substance on the left side always had an intensity of “fairly”; the intensity of the substance on the right side varied in a factorial design with the following levels: “very slightly,” “fairly,” and “very”. The intensity was also described in text next to each container (e.g., “This type of drink is fairly sweet”). The participant watched while the markers on both meters moved to indicate the intensity of the substance and then they clicked the mouse to proceed.

After the participant clicked the mouse, a third container appeared to the right of the original right-side container. This container had either a small, medium, or large amount of the substance in it. Participants were asked what the intensity of the property would be given this much of the substance. For example, “If I have a large amount of very heavy candy, what is its weight?” A black arrow appeared next to the marker on the meter on the right. Participants were asked to adjust the meter using the mouse so that it showed the intensity of the property for the specified amount of substance. When participants released the mouse an “OK” button appeared in the middle of the screen. Participants were told to press the “OK” button when they finished adjusting the meter and were informed that they could adjust the meter after they had released the marker the first time if they desired. After they clicked the “OK” button, the original right-side container disappeared. The marker on the right-side meter remained at the position the participant had selected. (This meter now indicated the judged intensity of the substance in the remaining right-side container.)

Next, the first part of the final question was displayed. The final question asked what the intensity of the property of the substance would be if both containers were combined. This question was displayed in three parts to minimize reading demands. Participants clicked the mouse to display the next part of the question. For example, the first part of the question might read, “If I take this large amount of very sweet drink. . . . ,” referring to the substance on the right. The second part would read, “. . .and combine it with this medium amount of fairly sweet drink. . . . ,” referring to the substance on the left. The third part would read, “. . .what is the sweetness of all the drink together?” When the third part of the question was displayed, the contents of the container on the right was added to the container on the left, although it was also still shown on the right for comparison. The original amount in the container on the left was shown by a red line on the container. Participants were asked to adjust the left-hand meter using the mouse so that it showed the intensity of the property for the combination of the two substances. The “OK” button appeared after the participant released the mouse. Again, participants were instructed to click the “OK” button when they were done
adjusting the marker. A “Ready” screen appeared after participants clicked the “OK” button. Clicking a button on the “Ready” screen started the next trial.

The experimenter explained each phase of the task to each participant during the first practice trial. Consistent with past research that used similar tasks with participants of comparable ages (e.g., Ahl et al., 1992; Moore et al., 1991), even the youngest participants appeared to have little difficulty with the task.

Materials

The containers were each 14.5 cm high and 4.5 cm wide, as measured on the surface of the computer screen. At the outset of the task, each container was filled with 4.5 cm of substance, as measured from the bottom of the container. The meter was 22 cm high and .75 cm wide. The marker on each meter was .5 cm high and 1.5 cm wide. The different intensities corresponded to different places on the meter: “very slightly” was 1.75 cm from the bottom of the meter, “fairly” was 6 cm from the bottom of the meter, and “very” was 9.5 cm from the bottom. The amount of the substance in the container on the right varied: “small” filled the container 2.5 cm from the bottom, “medium” filled it 4.5 cm from the bottom, and “large” filled it 8.0 cm. Participants working in the drink domain saw drawings of blue liquid in each container. Participants working in the candy domain saw drawings of a shiny green substance in each container. The candy looked like a hard, green substance, similar to rock candy.

Principle Scoring

Each principle specifies a different relationship between variables in the problem. Judgments about the problem are made by combining the available principles. To the extent that a participant represents a principle for a particular property, his or her judgments about that property should be consistent with that principle. Each participant’s judgment pattern was scored for consistency with each of three principles: range, monotonicity, and quantity. This procedure was followed for each of the three properties, weight, sweetness, and hemry, resulting in a set of three principle scores for each of the three properties. The principle scores measure the degree to which a participant instantiates that principle for the property. Past research has validated this methodology and these particular principles (Ahl et al., 1992; Dixon & Moore, 1996, 1997; Moore et al., 1991; Reed & Evans, 1987).

Range

The range principle specifies the relationship between the final intensity and the two intensities being combined and takes on different values for the additive and averaging models. Recall that, for the additive model, the final intensity must always be greater than either of the combined intensities; for the averaging model, the final intensity must always be between the two intensities. We scored each participant’s judgment pattern, such that high scores are consistent with the averaging model and low scores are consistent with the additive model; partici-
pants received 1 point for each judgment that was between the two intensities. Because another principle of the averaging model not considered here (Equal-Intensities-Equal) is involved when the two initial intensities are equal (i.e., “fairly” combined with “fairly”, these trials were not used in the scoring. The minimum possible score was 0; the maximum was 6.

Monotonicity

The monotonicity principle specifies the same relationship for both models; as the added intensity increases, the answer increases. We scored this principle by comparing the ordering of adjacent intensity judgments holding quantity constant. For example, a participant’s judgment when a “fairly” intense substance is added should be higher than his or her judgment when a “very slightly” intense substance is added (holding quantity constant). Each participant received 1 point for each correctly ordered pair of judgments within each added quantity. The minimum possible score was 0 and the maximum was 6.

Quantity

The quantity principle specifies the interaction between quantity and intensity. This principle takes on different values for both models. We scored this principle such that high scores were indicative of the additive model and scores near zero were indicative of the averaging model. The score was computed by comparing the ordering of adjacent quantity judgments within each level of intensity. For example, a participant’s judgment when a large amount of a fairly intense substance is added should be higher than his or her judgment when a small amount of a fairly intense substance is added. One point was given for each ordering consistent with the additive model; 1 point was deducted for each ordering that was inconsistent with the additive model. Nothing was done for ties (i.e., adjacent judgments being equal to each other). The minimum possible score was –6; the maximum was 6. Scores near zero are predicted by the averaging model, scores near 6 by the additive model. Scores near –6 would suggest that quantity was used subtractively, however, the vast majority of participants had scores between 0 and 6 (90% or more for all properties).

RESULTS

Overview

First, we present judgment patterns for each age group for the two familiar properties, weight and sweetness. Each age group’s judgments for weight and

---

3We initially created two scores each for the range and quantity principles, one to index consistency with the additive model and one to index consistency with the averaging model. For example, we created one score to indicate the degree to which each participant’s judgment pattern was consistent with the range principle specified by the additive model and a separate score for the range principle specified by the averaging model (and analogously for the quantity principle). However, examination of these scores showed them to be near perfect inverses of each other. Therefore, in order to simplify the presentation, we designed a single score for each principle. This score captures the conflicting additive and averaging model predictions.
sweetness showed a systematic pattern suggesting that, consistent with past research, they used intuitive models for combining these familiar properties. 

Next, we examine the pattern of judgments for combinations of hemry.4 Of par-

FIG. 4. The mean judged final intensity for weight plotted as a function of added intensity, with a separate curve for each added quantity. The vertical axis corresponds to the “weight meter,” measured in pixels. The dashed line shows the initial weight of the container on the left, which was held constant. Judgments made in the candy domain are shown in the left panels. Judgments made in the drink domain are shown in the right panels. The 5th-graders’ judgments are in the top panels, the 8th-graders’ in the middle, and the college students’ in the bottom panels.

sweetness showed a systematic pattern suggesting that, consistent with past research, they used intuitive models for combining these familiar properties. Next, we examine the pattern of judgments for combinations of hemry.4 Of par-

4Each participant made two judgments for each trial: a judgment about the change in intensity given a quantity change and a judgment about intensity when the two containers were combined. Usually, this first judgment is left implicit in mixture tasks. We chose to make it explicit to reduce the amount of integration the participant had to perform at one time. The analyses we report here use the second judgment (i.e., the judgment of the intensity when the two containers are combined) as the dependent variable. However, analysis of the first judgment yields very similar results, as does creating a composite score from the two judgments.
particular interest is whether these judgments also showed a systematic pattern, a result that would be consistent with the possibility that participants constructed an initial intuitive model for the novel property.

Next, the principle scores for each property are presented for each age group and domain. We then present evidence showing that, across age groups, participants constructed their initial model for the novel property by transferring principles from already-understood models. Finally, we use multiple regression to test the causal pathways predicted by the default model and domain hypotheses.
Judgment Patterns for Familiar Properties: Weight and Sweetness

Figure 4 shows the mean judgments for combinations of weight and Fig. 5 shows the mean judgments for combinations of sweetness. The vertical axis corresponds to the meter on which participants made their judgments, measured in pixels from the bottom of the meter. The dashed line at 175 indicates the initial intensity of the left-side container. In each figure, judgments for the candy domain are shown in the left-side panels and judgments for drink in the right-side panels. Each age group’s judgments for the familiar properties were analyzed in a $2 \times 2 \times 3 \times 3$ mixed ANOVA. All reported effects are significant at the .05 level.

Fifth grade. The top panels of Figs. 4 and 5 show the 5th-graders’ judgments for weight and sweetness in each domain. Overall, 5th-graders’ judgments showed significant effects of intensity, $F(2, 92) = 30.77$, and quantity, $F(2, 92) = 86.68$. The interaction between intensity and quantity was not significant, $F < 1$.

Surprisingly, the 5th-graders’ judgment patterns for weight and sweetness were not significantly different from one another. The main effect of property was not significant, $F < 1$, nor did the effects of intensity or quantity depend on which property, weight or sweetness, was being combined, $F < 1$, and $F(2, 92) = 1.03$, respectively. Similarly, the interaction between intensity and quantity did not depend on property, $F < 1$.

However, 5th-graders’ judgments showed effects of domain. As can be seen in the top panels of Figs. 4 and 5, for both weight and sweetness, judgments made in the candy domain were shifted up relative to judgments made in the drink domain, $F(1, 46) = 8.35$. Similarly, the effect of quantity was greater in the candy domain than in the drink domain, $F(2, 92) = 5.29$, although the effect of intensity did not depend on domain, $F < 1$. Both the higher judgments and increased effect of quantity in the candy domain are consistent with the idea that it was easier for the 5th-grade participants to apply an additive model to solid domains and an averaging model to liquid domains.

Eighth grade. The 8th-graders also showed significant overall effects of intensity, $F(2, 66) = 72.74$, and quantity, $F(2, 66) = 76.13$, as well as an intensity by quantity interaction, $F(4, 132) = 5.88$. However, unlike the 5th-graders, the 8th-grade participants (middle panels of Figs. 4 and 5) showed evidence of distinguishing between weight and sweetness. Their judgments for weight were shifted up relative to their judgments for sweetness, $F(1, 33) = 21.55$, and the effect of quantity was greater for judgments of weight than for judgments of sweetness, $F(2, 66) = 12.86$, although the effect of intensity was not, $F < 1$. The interaction between intensity and quantity did not depend on the property, $F(4, 132) = 2.25$, ns.

Also in contrast to the 5th-graders, the 8th-graders’ judgments were not significantly affected by the different domains. Their judgments in the candy domain were not shifted up compared to judgments in the drink domain, $F < 1$, nor were the effects of quantity, $F < 1$, or intensity, $F(2, 66) = 1.73$ ns, dependent upon domain.
College. The college participants’ judgments of familiar properties also showed overall effects of intensity, $F(2, 72) = 122.40$, and quantity, $F(2, 72) = 70.35$ (lower panels of Figs. 4 and 5). The interaction between intensity and quantity was significant, $F(4, 144) = 28.23$.

Similarly to the 8th-graders, college participants showed evidence of distinguishing between weight and sweetness. Their judgments for weight were shifted up compared to their judgments for sweetness, $F(1, 36) = 97.97$. Also, the effect of both quantity and intensity was greater for judgments of weight compared to those for

---

**FIG. 6.** The mean judged final intensity for the novel property, hemry, plotted as a function of added intensity, with a separate curve for each added quantity. The vertical axis corresponds to the “hemry meter,” measured in pixels. The dashed line shows the initial sweetness of the container on the left, which was held constant. Judgments made in the candy domain are shown in the left panels. Judgments made in the drink domain are shown in the right panels. The 5th-graders’ judgments are in the top panels, the 8th-graders’ in the middle, and the college students’ in the bottom panels.
sweetness, $F(2, 72) = 37.22, F(2, 72) = 5.53$, respectively. However, the interaction between intensity and quantity did not depend on the property, $F(4, 144) = 1.07$ ns.

Like the 5th-graders, college participants showed some evidence of the effect of domain. Although their judgments in the candy domain were not shifted up relative those in the drink domain, $F(1, 36) = 2.41$, ns, the effect of quantity was greater for judgments in the candy domain, $F(2, 72) = 3.19$.

**Judgments for the Novel Property: Hemry**

Figure 6 shows the mean judgments for the novel property, hemry. Judgments made in the candy domain are on the left side, and judgments made in the drink domain are on the right. Each age group’s judgments for combinations of hemry were analyzed in a 2 (Domain: Candy vs Drink) $\times$ 2 (Order: Weight-Sweetness vs Sweetness-Weight) $\times$ 3 (Intensity) $\times$ 3 (Quantity) mixed ANOVA. The order of the presentation of the familiar properties was included in the ANOVAs to test two hypotheses: (a) participants decided to “do what I did last time” when they encountered a novel property and (b) participants were primed by the last property they experienced. Both these hypotheses predict that judgments for hemry will be similar to the judgments made for the property that was presented immediately prior to hemry.

**Fifth grade.** The upper panels of Fig. 6 show the 5th-graders’ mean judgments for the novel property, hemry. As can be seen in the figure, the 5th-graders made systematic judgments about the novel property. That is, their judgments showed reliable effects of both intensity, $F(2, 92) = 24.48$, and quantity, $F(2, 92) = 36.73$, information. The interaction between intensity and quantity was also significant, $F(4, 184) = 3.76$.

As Fig. 6 suggests, 5th-graders’ judgments of the novel property differed significantly depending on the domain. Their judgment patterns in the candy domain are shifted up relative to their judgments in the drink domain, $F(1, 46) = 9.77$. Likewise, the effect of quantity is greater in the candy than the drink domain, $F(2, 92) = 6.01$. These effects are consistent with the hypothesis that the initial model participants generate depends on the domain. We explore more direct evidence for this hypothesis using the principle scores in a later section.

The order of the familiar properties (i.e., Weight-Sweetness vs Sweetness-Weight) had a significant main effect on 5th-graders’ hemry judgments, $F(1, 46) = 10.20$. However, the effect was the opposite of that predicted by both the “do what I did last time” and priming hypotheses. Participants who made judgments about weight immediately prior to making judgments for hemry made judgments that were shifted down rather than up. Order did not significantly interact with either quantity or intensity, $F$’s < 1, nor did the interaction between quantity and intensity depend on order, $F < 1$.

**Eight grade.** Like the 5th-graders, the 8th-grade participants made systematic judgments about the novel property. Their judgments, shown in the middle panels of Fig. 6, showed effects of both quantity, $F(2, 66) = 41.48$, and intensity, $F(2, 66) = 70.11$. There was also a significant interaction between intensity and quantity,
$F(4, 144) = 7.38$. However, unlike the 5th-graders, the 8th-graders’ judgments about the novel property were unaffected by the domain manipulation, $F < 1$, nor did domain interact with quantity, $F(2, 66) = 1.06$, or intensity, $F(2, 66) = 2.63$.

The order of the familiar properties did not have a significant main effect on the 8th-graders’ judgments of hemry, $F < 1$. Nor did order interact with intensity, $F(2, 66) = 1.51$, or quantity, $F(2, 66) = 1.25$. The interaction between quantity and intensity did not depend on order, $F(4, 132) = 1.88$.

College. The college participants’ judgments about the novel property, shown in the bottom panels of Fig. 6, were also systematic. Both intensity, $F(2, 72) = 93.43$, and quantity, $F(2, 72) = 20.21$, had significant effects. The interaction between intensity and quantity was also significant, $F(4, 144) = 9.43$. Like the 8th-grade participants, the college participants’ judgments were not affected by the domain manipulation. The main effect of domain was not significant, $F < 1$, nor did domain interact with intensity or quantity, $F's < 1$.

The order of the familiar properties did not have a significant main effect on college students’ judgments about hemry, $F < 1$. Similarly, order did not interact with either intensity or quantity, $F's < 1$, nor did the interaction between intensity and quantity depend on order, $F(1, 144) = 1.25$.

In summary, 5th-graders, 8th-graders, and college participants made sets of orderly judgments for combinations of hemry across randomly ordered trials. One reasonable explanation of these results is that participants constructed and used an intuitive model of how hemry combines. The mean judgment patterns for hemry also suggest that domain affected the 5th-graders’ initial models, but not the initial models constructed by the older participants. None of the age groups showed the effects of order of the familiar properties predicted by the “do what I did last time” and priming hypotheses. Only the 5th-grade students showed an effect of order and this effect was not in the predicted direction.

Analysis of Principle Scores

Table 2 shows the means and standard deviations for the principle scores for each property with a separate column for each domain and a separate row for each grade. Means and standard deviations for the monotonicity principle scores are in the two leftmost columns. Means and standard deviations for the range principle scores are in the two middle columns, and those for quantity are in the two rightmost columns. The top panel shows the principle scores for the weight property. The middle panel shows the principle scores for the sweetness property and the bottom panel for the novel property, hemry.

Table 2 reveals that there was considerable variability in participants’ use of the principles. As can be seen by examining the standard deviations, there was substantial within-age-group variability in the familiar and novel property principle scores. Even the college participants did not uniformly use one set of principles for either of the familiar properties. While this variability makes the interpretation of the mean judgment patterns less clear, it allows for examination of how individual differences in participants’ models for familiar properties are related to the models.

they construct for the novel property and whether this relationship changes developmentally. In particular, this variability allows us to test our first hypothesis, that the initial model is formed by transferring principles from familiar properties.

**Transferring principles from familiar to novel properties.** If participants transfer principles from the familiar properties to their model for the novel property, the correlation matrix between the principle scores for the familiar and novel properties should have a particular structure. Both isomorphic and nonisomorphic principles will be correlated. Isomorphic principles specify analogous structural relationships (e.g., range for weight and hemry are isomorphic). Nonisomorphic principles specify different structural relationships (e.g., quantity for weight and range for hemry are nonisomorphic). Isomorphic principles will be correlated because the principle itself is transferred to the initial model for the new property. For example, if the range principle from the model for sweetness (range-sweetness) is transferred to the model for hemry, then the range-sweetness score will be correlated with the range principle score for hemry (range-hemry).

Nonisomorphic principles will also be correlated, but only because principles tend to develop together within a property. For example, children develop the monotonicity principle for sweetness while they are also developing the range principle for sweetness (Ahl et al., 1992; Moore et al., 1991). Therefore, if range-sweetness is transferred to the model for hemry, range-hemry will be correlated with monotonicity-sweetness because of the correlation between range-sweetness and monotonicity-sweetness. However, the correlation between monotonicity-sweetness and range-hemry will be entirely mediated by range-sweetness.
This leads to a crucial prediction for the principle transfer hypothesis. Only an isomorphic principle for a familiar property should ever explain unique variance in the principle score for a novel property. Nonisomorphic principles, while correlated, should never explain unique variance in principle scores for hemry. For example, the semipartial correlation between range-sweetness and range-hemry should be significant, but the semipartial between monotonicity-sweetness and range-hemry should not be.

Table 3 shows the bivariate correlations between the principle scores for each familiar property, sweetness and weight, and the novel property, hemry, separately for each grade. The top panel shows the results for the 5th grade, the middle panel shows results for the 8th grade, and the lower panel shows results for college participants. The three columns on the left show the correlations between principle scores (monotonicity, range, and quantity) for weight and hemry. The three columns on the right show the correlations between the principle scores for sweetness and hemry. Underlined correlations are significant. Boldface correlations also have a significant semipartial correlation when the contributions of the other principles are partialed out. Notice that, across all grades, only correlations on the diagonals, between isomorphic principles, have significant semipartial correlations. That is, only isomorphic principles of familiar properties explain unique variance in the hemry principle scores. For example, for the 8th-graders the correlation and the semipartial correlation between quantity-sweetness and quantity-hemry are significant. The correlation between monotonicity-sweetness and quantity-hemry is significant, but the semipartial correlation is not.
This pattern of results supports the hypothesis that, across age groups, participants are transferring the principles from the familiar properties to the initial model for the novel property. The 5th- and 8th-graders appear to transfer principles from both weight and sweetness to hemry. The college participants only appear to transfer principles from sweetness. The college participants’ lack of transfer from their model for weight is potentially interesting because it suggests that the averaging model is predominant as their initial model for combination. The principle transfer hypothesis is, of course, consistent with transfer from either both models (i.e., the 5th and 8th graders) or a single model (college students). The hypothesis specifies how transfer occurs, across isomorphic principles, but does not specify which models people transfer. The transfer process appears to operate similarly across this fairly wide age range, despite the very apparent differences in the quality of the familiar models across grades.

Given the strong evidence for the transfer hypothesis from the pattern of correlations, we constructed causal models for transferring principles to the novel property, hemry. These models allow us to examine the remaining two hypotheses: (a) that one or more age groups used a default model and (b) that domain affected transfer.

Causal Models of Principle Transfer

For each age group, we estimated a set of three causal models, one for each principle. The models estimate the degree of transfer of the principles from the familiar properties, sweetness and weight, to the novel property, hemry. The models also estimate the potential effects of domain and the order of presentation of the familiar properties. The principle scores for hemry are the outcome variables. Figure 7 shows the potential causal pathways. (Because order did not show the predicted effects on transfer, the pathways for order are omitted to simplify the figure). Transfer is predicted to be constrained by the degree to which the principle is used for the familiar property. Therefore, individual differences in using the principle for the familiar property should affect transfer to the novel property. For example, a child whose range principle for sweetness is not fully developed could only transfer that partially developed principle to the model for hemry.

We hypothesized that the type of domain (i.e., whether judgments were made about properties of drink or candy) could affect transfer both directly and indirectly. The indirect effect was hypothesized to work through the familiar properties. If participants can better instantiate the principles for a familiar property in one domain, then transfer to the novel property should be facilitated in that domain as well. For example, if the principles for weight were easier to instantiate in the candy domain, the principles of the model for weight would be more available for transfer to the initial model than in the drink domain.

Domain could also potentially have a direct effect on transfer. If the additive model is more prevalent in solid domains and the averaging model is more prevalent in liquid domains, working within a particular domain may activate the more
prevalent model. The principles from the activated model would be available for transfer to the initial model for the novel property.

The order of the familiar properties (i.e., Weight-Sweetness vs Sweetness-Weight) might affect transfer because the model for the property presented immediately prior to hemry would be primed. For example, if a participant first made judgments about weight then judgments about sweetness, the averaging model would be primed when the participant encountered hemry. Therefore, principles from the averaging model would be more likely to transfer. The hypothesis that participants might simply continue making judgments using the same principles as they did for the last familiar property (i.e., “do what I did last time”) makes the same prediction. However, order was not significantly correlated with any of the hemry principle scores for any grade, with the exception of range for the 5th-graders, $r (48) = .35$, and this effect is in the wrong direction. Fifth graders’ range scores for hemry tend to be larger, indicative of the averaging model, when immediately preceded by weight. One explanation of this effect is that the younger children may find it difficult to switch among

**FIG. 7.** Causal models predicting the transfer of principles from sweetness and weight to hemry for each grade. In each model, domain refers to whether the judgments were about a property of drink or candy. Black arrows denote significant pathways; gray arrows denote nonsignificant pathways. The numbers on the significant pathways are semipartial correlations.
models, so that at least some aspects of the model used first persist across the other properties.

The top panel of Fig. 7 shows the causal pathways for each principle for the 5th grade. Gray lines indicate nonsignificant relationships and black lines indicate significant relationships. The semipartial correlation between the predictor variable and the principle score for hemry is shown for each significant pathway.

As can be seen in the figure, the 5th-grader’s principle scores for both sweetness and weight explained unique variance in their principle scores for hemry. For sweetness, the semipartial correlations were \( sr'(45) = .27, .25, \) and \(.25, \) for monotonicity range, and quantity, respectively. For weight, the semipartial correlations were \( sr'(45) = .41, .37, \) and \(.30, \) for monotonicity, range, and quantity. The 5th-graders appear to transfer principles from both familiar properties to the novel property.

Domain did not significantly affect any of the principle scores for the familiar properties. The sweetness principle scores were unrelated to domain, \( r'(48) = -.25, -.16, \) and \(.24, \) as were the weight principle scores, \( r'(48) = -.01, -.17, \) and \(.15, \) for monotonicity, range, and quantity, respectively. This rules out the possibility of domain indirectly affecting the hemry principle scores through the familiar properties.

However, domain did appear to have a direct effect on the hemry principle scores. Domain explained significant unique variance in the hemry principle scores, \( sr'(46) = -.22 \) and \(.30, \) for range and quantity, respectively. The 5th-graders’ range and quantity principles for hemry were more appropriate for the averaging model when hemry was a property of drink and more appropriate for the additive model when hemry was a property of candy.

As can be seen in the lower left panel of Fig. 7, the 8th-graders’ principle scores for the familiar properties also explained unique variance in the hemry principle scores. For sweetness the semipartial correlations were \( sr'(32) = .44, .42, \) and \(.30, \) for monotonicity, range, and quantity, respectively. For weight the semipartial correlations for monotonicity and range were significant, \( sr'(32) = .38 \) and \(.42, \) but the semipartial correlation for quantity was not, \( sr(32) = .18. \) In general, the 8th-graders transferred principles from both familiar properties to the novel property.

For the 8th-graders, domain did not significantly affect the principle scores for the familiar properties, largest absolute value of coefficient, \( r(35) = .17, \) nor did domain predict variance in the hemry principle scores, largest absolute value of the correlations, \( r(35) = .18. \)

The results for the college participants are shown in the lower right panel of Fig. 7. College participants principle scores for sweetness explained unique variance in the hemry principle scores, \( sr'(35) = .34, .65, \) and \(.42, \) for monotonicity, range, and quantity, respectively. The principle scores for weight did not explain unique variance in the hemry principle scores, \( sr'(35) = .22, .20, \) and \(.07, \) for monotonicity, range, and quantity.

Domain did not significantly affect college participants’ principle scores for either of the familiar properties, largest absolute coefficient value, \( r(38) = .22, \) nor did domain explain variance in the hemry principle scores, largest absolute value of the correlations, \( r(38) = .21. \)
Because the semipartial correlations provide very conservative estimates of the effects, it is worth noting that the majority of the models explain a respectable proportion of the variance in the hemry principle scores. For the 5th-graders, the models accounted for 30, 50, and 34%, of the variance in the hemry monotonicity, range, and quantity principle scores, respectively. For the 8th-graders the models accounted for 49, 48, and 19%, of the variance in these scores. The models for the college students accounted for 31, 57, and 21%, of the variance in their hemry principle scores for monotonicity, range, and quantity, respectively.

**DISCUSSION**

Supporting the assumption that people generate initial models when they encounter a novel property, all age groups made sensible and systematic judgments about hemry. That is, their judgment patterns showed predictable effects of both intensity and quantity. Analysis of the principle scores showed that there were systematic relationships between the judgment patterns for the familiar and novel properties, indicating that responses to the novel property were themselves systematic. Both the systematic effects of intensity and quantity and the systematic relationships between the familiar and novel properties are consistent the assumption that people generated an intuitive for the novel property.

The pattern of correlations between the familiar and novel principle scores supports the hypothesis that individual principles are transferred from the familiar properties to the novel property. This provides some insight into how initial models are constructed. The models people first generate for novel properties appear to consist of relationships borrowed from already-understood properties. This is consistent with research on analogy which shows that relationships between very dissimilar domains can, under certain conditions, be transferred (Gick & Holyoak, 1983; Holyoak & Thagard, 1997; Ross & Kilbane, 1997, Spellman & Holyoak, 1996). Note that the additive and averaging models are both good analogical solutions for the novel property, according to theories of analogical mapping. For example, both of these solutions have very good structural consistency, relational focus, and systematicity (see Gentner, 1983; Gentner & Markman, 1997). In the present study, we showed the transfer of specific principles or structural relationships by measuring the aspects of performance predicted by each principle. The transfer results suggest that principles are the substrate used to construct initial models. The transfer of principles from the familiar properties to the novel property also implies that children’s initial models are constrained by the quality of the intuitive models they already have in their repertoires. Therefore, one important implication of the present study is that a child’s prior understanding of properties will strongly influence the initial models he or she uses (i.e., the starting point in the model space) and subsequently the knowledge he or she acquires.

We hypothesized that younger participants would use the additive model as a default whenever a novel property is encountered. The evidence against this hypothesis is quite strong for all age groups. Consider the 5th-graders, who were predicted to show this effect most pervasively. Although their mean judgment pat-
tern for the novel property was not completely consistent with the normative averaging model for either domain, the principle scores for the weight and sweetness properties explained unique variance in principles scores for the novel property. This shows that some 5th-graders were transferring principles from their averaging model, the model appropriate for sweetness; other participants transferred principles from the additive model, the model appropriate for weight. Note that the hypothesis that some participants were simply applying the additive model to both properties, sweetness and weight, cannot explain the unique variance that was contributed by principles from each property. If the same model was used for sweetness and weight, the semipartial correlations would approach zero. A very similar pattern of results was observed for the 8th-graders.

Surprisingly, the college participants did not show any evidence of transferring principles from the additive model. Rather, only principle scores for sweetness, the averaging model, explained unique variance in the principle scores for the novel property. Especially strong evidence that transfer is only from the averaging model comes from the monotonicity principle. Recall that both the additive and averaging models have the same value for this principle; that is, both produce judgment patterns that are monotonically increasing. However, only individual differences in the monotonicity score for sweetness predict unique variance in the monotonicity score for the novel property.

Taken together, these results provide strong evidence against the hypothesis that the additive model is used as the default for new properties. Although there appears to be a developmental shift from the additive to averaging models for the novel property, even the 5th-graders sometimes applied their averaging model principles to the novel property. The finding that even the youngest children do not have a default model for novel properties suggests that the initial models they generate may be tailored to the new property. That is, in addition to making changes in the model as they accrue evidence about the new property, children may also use regularities in their environment to construct a “best bet” initial model. We hypothesized that one way children might do this is through the relationship between domains and models.

According to this hypothesis, the domain would help constrain the initial models people generate for novel properties. Because liquids tend to be distinguished by differences in concentrations, and solids tend to be distinguished by differences in extensive properties, we predicted the averaging model would be used for novel properties in the liquid domain and the additive model would be used for the solid domain. Domain had a direct effect on 5th-graders’ use of the range and quantity principles. As predicted, their principle scores indicated greater use of the averaging model for the drink domain and greater use of the additive model for the candy domain. This effect was statistically independent of their use of the range and quantity principles for the familiar properties. This suggests that domain is not simply affecting access to the models. For example, if the averaging model was easier to access in the liquid domain, we would have observed this effect for both the familiar and novel properties. However, domain did not affect
the range or quantity principles for the familiar properties; rather, the effect on the initial model appeared to be direct. One explanation of this effect is that when children encounter a novel property within a particular domain, they search for familiar properties of that domain. The most pervasive model among these familiar properties is used as the initial model for the new property.

Neither the 8th-graders nor the college participants showed an effect of domain, either directly or indirectly. This suggests that, with development, older children and adults begin to separate properties from the domains in which those properties were learned. One plausible hypothesis is that the developmental shift from domain specificity to domain independence results from the interaction between the structure of the physical world and the process of knowledge acquisition. Just as the structure of English verb tenses results in overgeneralizing regularities at particular points in language acquisition (e.g., Plunkett & Marchman, 1993), the structure of the physical world may also result in overextending regularities during knowledge acquisition. Younger children’s use of domain in generating their initial models may reflect, that in their early experience, particular domains are associated with particular models. Older children’s indifference to domain may reflect that, with greater experience, domains no longer accurately predict models. The developmental change in children’s use of the domain to generate initial models would stem from an increasingly weak relationship between domains and models as children learn additional properties. Regardless of whether this explanation is accurate, the current study adds to previous work showing that domain appears to play an important, but developmentally changing, role in children’s inductive processes. For example, Mandler and McDonough (1998) found that both 14- and 20-month-old children used domain information in inductive generalizations but what children view as constituting a domain changes with development. The younger children appeared to define domains very broadly, while the older children’s domains were nearer to adult domain categories (see also Springer & Keil, 1991).

It is worth noting that neither the priming nor the “do what I did last time” hypothesis was supported by the data. The order of presentation of the familiar properties did not affect the judgments for hemry as predicted by these hypotheses; indeed the only significant effects of order conflict with these hypotheses. This pattern of results is consistent with other work from our laboratory, in which we investigated the effects of order more completely (Dixon, 1997). Even when the order of presentation of the novel property and the familiar properties is completely counterbalanced, order does not significantly affect transfer. In this additional study, the novel property and the familiar properties were presented equally often in all order positions. The same pattern of correlations indicative of transfer was observed across presentation orders and order did not significantly affect the extent of transfer.

Both the domain and children’s understanding of familiar properties constrain the construction of initial models, but our results also suggest that these constraints change with development. As Sophian (1997) cogently argued, con-
straints must change as a result of development if they are to be useful for development. The core of her argument is that the constraints that children use to limit the number of potential inductive inferences (or models) would eventually prohibit further development unless the constraints themselves underwent developmental changes. For example, in word learning, the whole-object constraint (Markman, 1990), which initially limits the interpretation of word meanings, would eventually interfere with learning names for object parts.

Consistent with Sophian’s position, young children’s use of domain allows them to limit the potential models they construct for a new property, but if domain remained a strong constraint, less prevalent models would be very difficult to discover in heavily dominated domains. Similarly, as children’s current models for familiar properties improve with development, the effect of their current models as a constraint changes. Younger children’s relatively undifferentiated models for familiar properties result in very similar initial models, regardless of which property is used as the base. Older children’s well-differentiated models for familiar properties result in very different initial models, depending on which property is used as the base. Therefore, the effect of children’s familiar models as a constraint on their initial models changes radically with development.

An important question raised by the centrality of transfer in this process is how children arrive at their original base models. As mentioned in the introduction, at some point during development children must construct each type of model (e.g., the averaging model) for the very first time. One possibility is that privileged properties, properties for which the appropriate models are easily acquired because of cognitive and/or perceptual specializations, may be the original base models for transfer. According to this view, children develop their first instance of each type of intuitive model for privileged properties. When children later encounter nonprivileged properties, they transfer their models for privileged properties to their initial models for the nonprivileged properties. This position is appealing because it proposes that children use specialized cognitive and perceptual processes to solve the difficult logical problem of constructing models from scratch and use more general processes, similar to analogical transfer, to extend these models to the broader array of nonprivileged properties.

REFERENCES


Halford, G. S., Brown, C. A., & Thompson, R. M. (1986). Children’s concepts of volume and flota-


