The Role of Language, Appearance, and Culture in Children’s Social Category-Based Induction

Gil Diesendruck and Heidi haLevi

Bar-Ilan University

Four studies examined whether Israeli 5-year-olds (N = 88) and adults (N = 48) drew inferences about psychological properties based on a character’s social category, personality trait, or physical appearance trait. Study 1 revealed that while children drew inferences mostly by social category, adults did it by personality trait. Study 2 showed that the children’s pattern was not due to how the categorical information was conveyed. Studies 3 and 4 demonstrated that for kindergarteners, labels, not appearances, are determinant of the inductive potential of social categories. Studies indicated that “Jew” and “Arab” were the most inductively powerful social categories for both children and adults. The results carry implications for the roles of language, appearances, and culture in the conceptualization of “human kinds.”

Categories are extremely useful cognitive tools for making sense of the world. As information-rich cognitive representations, categories save us from treating each individual object as an entirely novel entity. In fact, it has been argued that categories are somewhat like “theories” (Murphy & Medin, 1985). The argument is that people believe that there are fundamental and specific reasons for why things belong to the same category, and these beliefs license inferences or predictions about the characteristics of novel category members. Something is a “lion” not simply because of the way it looks. Rather, we believe that lions have a specific genetic makeup, which generates certain relatively stable and universal physical characteristics and behavioral dispositions. Consequently, by recognizing that an animal I see is a “lion,” I can infer what it might want from me, how fast it can run, and how I should react.

The notion of categories as rich sources of induction is also related to the idea that people might have a belief that certain categories have essences that define what they are and what properties they have—a belief Medin and Ortony (1989) termed “psychological essentialism.” A substantial body of literature indicates that young children treat animal categories in a way consistent with psychological essentialism (see Gelman, 2003, for a review; but see Strevens, 2000, for an alternative interpretation of this literature). In particular, children seem to believe that animal categories have innate potential and are resistant to superficial transformation (Gelman & Wellman, 1991), are defined by internal nonobvious properties (Diesendruck, Gelman, & Lebowitz, 1998), and have rich inductive potential (Gelman & Markman, 1986, 1987).

Lately, researchers have asked to what extent children and adults hold similar beliefs about categories from other domains, especially categories of human groups (Hirschfeld & Gelman, 1994). Studies have found that people seem to reason about various human categories similarly to how they reason about animal categories (e.g., Astuti, Solomon, & Carey, 2004; Gil-White, 2001, on ethnicity; Heyman & Gelman, 1998, on personality traits; Hirschfeld, 1996, on race; Taylor, 1996, on gender). Findings like these have led some researchers to argue that people “essentialize” human categories, treating them as natural kinds (Hirschfeld, 1996).

These similarities notwithstanding, a crucial difference between animal and human categories considers the role of culture in categorization. Animal categories can be considered natural kinds in that they have clearly marked and universally recognized morphological and behavioral distinctions (Atran, 1990). Accordingly, children and adults from various cultures seem to hold similar beliefs about animal categories (Astuti et al., 2004; Atran, Estin, Coley, & Medin, 1997; Diesendruck, 2001; Sousa, Atran, &
Medin, 2002; Walker, 1999). In contrast, human categories seem to lack the natural kind characteristics of animals just mentioned (Rothbart & Taylor, 1992). Accordingly, cultures differ in terms of how they conceive of similar human categories, and which human categories they essentialize (cf., Astuti et al., 2004, in Madagascar; Gil-White, 2001, in Mongolia; Hirschfeld, 1996, in the United States and France; Mahalingan, 2003, in India). Given this principled difference between natural and social categories, and the findings of cross-cultural variation in the categorization of human kinds, two questions that have become the focus of much debate are which human kinds are essentialized, and what factors in the social environment contribute to this process of essentialization. Answers to these questions can help elucidate why human kinds are essentialized.

The goal of the present studies was to address these questions by investigating the relative inductive potential of different human categories within a given culture. First, we were interested in comparing the relative inductive potential of personality traits vis-à-vis social categories. The notion of human categories, or kinds, is often utilized to refer to both personality traits and social categories. However, the two seem to differ in important ways. To name just a few, most cultural discourse (e.g., the census) revolves around social categories rather than personality traits (Rothbart & Taylor, 1992), and stereotypes or ingroup–outgroup biases are most commonly directed toward social categories (Hilton & von Hippel, 1996). One of our goals was to investigate whether these differences between social categories and personality traits carry conceptual implications for both children and adults.

There are theoretical reasons to believe that they do. On the one hand, personality traits—especially in Western countries—are arguably the primary categorical distinction by which adults draw psychological inferences (Markus & Kitayama, 1991). Personality traits are treated as causal factors capable of explaining people’s behavioral tendencies and preferences (Yuill, 1992) and, in fact, even preschoolers rely on personality traits to draw psychological inferences (Heyman & Gelman, 2000). In other words, personality traits have a number of the markers of essentialist thinking (Gelman, 1992). On the other hand, certain characteristics of social categories (such as ethnicities and races) may make them especially salient for essentializing. Hirschfeld (1996) argues that children intuitively essentialize intrinsic human kinds, and that cultural discourse about social categories contributes to this conceptualization. Alternatively, Gil-White (2001) argues that essentialized human kinds are those in which members marry within the kind, in which kind membership is passed on by inheritance, and which have some distinctive physical markers (even if superficial ones such as clothing)—properties associated with certain social categories (e.g., ethnicities) but not with personality traits. These latter arguments favor the hypothesis that social categories should be more strongly essentialized than personality traits.

Second, we were interested in assessing the relative contribution of physical similarity and common labeling to the inductive potential of human categories. This interest is again grounded on a number of theoretical issues. If we were to find differences in the inductive potential of personality traits versus social categories, we wanted to be able to establish to what extent this difference would be due to the presence of physical distinctions in social categories. Hirschfeld (1996) argues that the essentialization of a specific human kind (e.g., race) derives first and foremost from the cultural discourse about the kind—not from its physical distinctiveness. In a similar vein, but from a different theoretical perspective, Sperber (1996) claims that language is one of the main factors driving the essentialization of human kinds. Others have claimed that physical markers substantially contribute to the treatment of social categories as if they were animal species (Atran, 1990; Gil-White, 2001). A systematic manipulation of these factors could thus help elucidate this issue.

Third and finally, we were interested in comparing the relative inductive potential of distinct social categories (e.g., ethnicity, religiosity), salient in a particular cultural milieu (Israeli secular society), to young children. Specifically, given Hirschfeld’s (1996) emphasis on the role of cultural discourse in social essentialism, we were interested in seeing whether young children would be sensitive to the same social categories as the adults in their culture. We targeted kindergarten children because while, on the one hand, they already seem to have stereotypes regarding certain social categories (in particular ethnicity, Bar-Tal, 1996; Teichman, 2001), these stereotypes seem to grow stronger as they enter into school and get more exposed to the cultural discourse about social groups (Al-Haj, 2002; Bar-Tal, 1998). We were thus interested in assessing children’s social essentialism before their more massive exposure to institutionalized cultural discourse. Moreover, kindergarten children are the youngest group for which evidence of social essentialism has been consistently reported (Gelman, 2003).

The methodology chosen to address these questions was one originally used by Gelman and
Markman (1986, 1987) in their studies of the inductive potential of animal categories, and adapted by Heyman and Gelman (2000) to study personality traits. Arguably, the inductive potential of categories is not only one of the central components of essentialism (Gelman, 2003), but is also a characteristic of categories with important practical implications, given its role in affecting how people interact based on assessments of others’ category membership (Yzerbyt, Judd, & Corneille, 2004). In Heyman and Gelman’s studies, an experimenter showed 5-year-old children two characters and told them about the different psychological properties they had (e.g., one liked to play “tibbits” whereas the other “jimjam’’). A third character was then presented, who was physically similar to one of the initial characters, but was described as having the same personality trait as the other initial character (e.g., they were both “nice’’). Children were then asked which psychological property the third character had. Analogous to Gelman and Markman’s findings with animals, Heyman and Gelman found that children extended psychological properties based on common personality trait rather than physical similarity. Using this methodology, Study 1 directly addresses the first and third goals described above. Studies 2, 3, and 4 help clarify the conclusions from Study 1, rule out alternative interpretations, and address the second goal.

**Study 1**

**Method**

**Participants.** Twenty-four adults (M = 22 years, range = 19–28 years) and 24 children (M = 5 years 7 months, range = 4 years 2 months to 6 years 4 months) participated in this study. There were 17 women and 7 men among the adults, and 13 girls and 11 boys among the children. All adults were university students who either volunteered or received course credit for participating. Children were recruited from kindergartens in the cities surrounding the university. All children came from middle-class, secular, Jewish families. Only children with signed parental permission participated in the study.

**Design.** “Age group” (children vs. adults) was the only between-subjects factor. The study consisted of eight trials, each presenting children with the possibility to make an inference based on either the social category membership or the personality trait of the target character.

**Materials.** In each trial, children were shown a triad of line-drawn lightly colored human characters: a target character and two test characters. One of the test characters was presented to children as being of the same social category as the target but as having a different personality trait (social category match). This character resembled the target physically only in terms of social category markers (e.g., “religious” male characters wore a yarmulke and other traditional clothing attire, “rich” characters wore more formal clothing and had a well-combed hairdo). The other test character was presented to children as having the same personality trait as the target but as belonging to a different social category (personality trait match). This character did not resemble the target physically (see Figure 1 for an example).

In order to emphasize the social category similarity, other physical characteristics of the characters were made irrelevant, either by equating them (e.g., same height) or by randomizing them (e.g., hair color). Moreover, for this same reason characters in each triad were of the same gender (except when “gender” was the social category tested). Three character triads were male and three were female.

**Procedure.** A female experimenter tested participants individually. Adults were tested in a university laboratory, whereas children were tested in a quiet area of their kindergarten.

The procedure started with the experimenter showing participants one of the eight picture triads. While concealing the target character, the experimenter pointed to each of the test characters in the triad, and told participants each character’s social category and personality trait. The experimenter then revealed the target character, pointed at it, aligned it with the test characters, and told the participants its social category and personality trait information, explicitly noting the similarities to each of the test characters. The experimenter next told participants about a novel property each of the test characters had. Finally, the experimenter reminded children about the target character’s social category and personality trait, and asked children to infer the target’s novel property. Figure 1 depicts a schematic example of the procedure.

The four social categories used were selected for their salience in Israeli culture and for their presumed familiarity to kindergarteners. They were “gender” (boy/girl), “social status” (rich/poor), “religiosity” (religious/secular), and “ethnicity” (Jewish/Arabic). The four personality traits used were selected for their reported familiarity to children of this age (Heyman & Gelman, 2000). They were “niceness” (nice/not nice), “shyness” (shy/friendly), “creativity” (creative/not creative), and “activeness” (active/quiet). Each social category pair
and each personality trait pair appeared twice during the procedure. All possible pairings of social category with personality trait were used across the sample. In other words, for each child, each social category was paired with different personality traits. The pairings were randomly assigned to participants, and they appeared the same number of times across the entire sample. The order of presentation of the pairings within any given participant was also random. The right–left placement of the test characters and consequently the order of presentation of the test characters—was counterbalanced within subjects. Also counterbalanced within subjects was the order of presentation of the social category and personality trait information about each test character.

The 16 novel properties were similar to those used in previous work addressing preschoolers’ inferences (Heyman & Gelman, 2000). All properties were novel; therefore children could not base their answers on stigmas tying particular properties to categories. They had to do with preferences, desires, and behaviors. They were “likes to play with ‘Bilan/Uma,’” “likes to do ‘pipo/zavo’ to his friend,” “spends a lot of time at the ‘roban/kore’an,’” “likes to ‘darin/kanken’ with his friend,” “wants to be a ‘mashitz/narim’ when he grows up,” “likes to play ‘zaber/zigo,’” “likes to pretend to be a ‘palkan/talan,’” and “listens a lot to ‘mishinit/randit’ music.”

Participants’ responses were scored as “1” if they inferred that the target character had the same property as the social category match, and as a “0” if they inferred that the target character had the same property as the personality trait match. Thus, a score of “8” indicated that a participant drew all his or her inferences based on social category, whereas a score of “0” indicated that a participant drew all his or her inferences based on personality trait similarity.

Results and Discussion

The main dependent measure used in the analyses was the number of trials, out of eight, in which participants answered that the target character had the same property as the social category match.

In order to compare children and adults’ inferential patterns, we first conducted an analysis of variance (ANOVA) with age group and gender as between-subjects factors. This analysis revealed a significant effect of age group, with children making more social category based inferences than adults, $F(1,44) = 18.20, \ p < .001, \ \eta^2 = .293$ (see Table 1).
Neither the main effect of gender nor the interaction between gender and age group was significant. One-sample t tests were conducted separately for children and adults in order to compare their pattern of responses to chance (chance = .4). These analyses revealed that children made social category-based inferences more often than expected by chance, \( t(23) = 2.64, p < .05 \), whereas adults made social category-based inferences less often than expected by chance, \( t(23) = -4.71, p < .001 \). That is, adults systematically drew inferences based on personality traits.

We conducted nonparametric analyses to assess the different strategies participants employed. Using the same classificatory system as the one used by Heyman and Gelman (2000), we classified participants as “personality trait reasoners” if they drew inferences based on personality traits on six or more of the trials, as “social category reasoners” if they drew inferences based on social category on six or more of the trials, and as “inconsistent reasoners” if they manifested a different distribution. Table 2 presents this distribution. Replicating the findings from the ANOVA, we found that there were more social category reasoners among children than among adults, and the opposite was the case in terms of personality trait reasoners, Fisher’s exact test: \( p < .001 \). Further analyses revealed that both the children, \( \chi^2(1, N = 24) = 11.04, \text{exact} \ p < .01 \), and the adult distribution, \( \chi^2(1, N = 24) = 7.03, \text{exact} \ p < .05 \), significantly differed from the distribution expected by chance.

Our next analyses looked at potential differences among social categories. For this purpose, we conducted separate t tests for each of the social categories, on the frequency with which participants inferred based on social category membership. Recall that for all participants, every social category appeared in two trials out of the eight, each time paired against a different personality trait. These analyses revealed that for only two of the social categories did children make social category-based inferences significantly more than expected by chance (chance = 1); namely, on ethnicity (\( M = 1.4, SD = 0.7 \), \( t(23) = 2.58, p < .05 \), and social status (\( M = 1.5, SD = 0.6 \), \( t(23) = 3.82, p < .005 \). Interestingly, these were the only two social categories for which adults made social category-based inferences at chance. That is, for both gender (\( M = 0.7, SD = 0.5 \), \( t(23) = -3.39, p < .005 \), and religiosity (\( M = 0.7, SD = 0.7 \), \( t(23) = -2.07, p = .05 \), adults made social category based inferences less than expected by chance (i.e., they made the inferences based on personality trait). We also conducted complementary analyses looking at the number of times participants drew inferences based on social category for each personality trait (the complementary frequencies reflect the number of inferences based on the specific personality trait). The only personality trait selected by children significantly less than expected by chance was shyness (\( M = 1.3, SD = 0.7 \), \( t(23) = 2.07, p = .05 \). Adults selected activity (\( M = 0.7, SD = 0.7 \), \( t(23) = -2.07, p = .05 \), and niceness (\( M = 0.7, SD = 0.8 \), \( t(23) = -2.15, p < .05 \), significantly more than expected by chance, and creativeness (\( M = 0.8, SD = 0.6 \) marginally more than expected by chance, \( t(23) = -2.02, p = .056 \).

These findings indicate that despite the general difference between children and adults in their preferred source of inferences—children rely on social category, adults on personality traits—children and adults seem to agree on what are the most inductively powerful social categories among the four assessed here. Ethnicity and social status were the categories in which children’s tendency to draw inferences based on social category manifested most clearly, and they were the categories that most strongly moderated adults’ tendency to draw inferences based on personality traits. The correspondence between children and adults regarding the inductive potential of individual personality traits was not as clear-cut.
These latter analyses also reinforce the importance of presenting participants with a situation in which they had to decide between two plausible candidates (i.e., personality traits and social categories). This situation minimized the chance of getting ceiling performance (e.g., participants consistently inferring by social category), and thus increased the potential for variance among the social categories. The analyses also reinforce the importance of asking participants to draw inferences about novel properties that were not stereotypically associated with any particular social category. It is plausible that children are already aware that specific properties (e.g., dietary habits) are especially relevant to certain social categories (e.g., religiosity).

Taken together, the findings of Study 1 imply that, for children, social categories are more strongly essentialized human kinds than are personality traits. The findings are consistent with Hirschfeld’s (1996) argument about the importance of the cultural discourse in pointing to children essentialized human kinds, and especially with Gil-White’s (2001) account for why ethnicities are essentialized.

Study 2

An alternative explanation for the difference between children and adults’ reliance on social categories found in Study 1 has to do with a methodological issue. Namely, in Study 1, similarity in terms of two characters’ personality trait was conveyed only verbally. In turn, similarity in terms of social category membership was manifested both verbally and visually (i.e., the social category match shared some physical characteristics with the target character). Possibly, children’s tendency to select a social category match instead of a personality trait match simply reflected this asymmetry in the kinds of input they received about each of the test characters.

The goal of Study 2 was to address this possibility. For this purpose, Study 2 contrasted purely physical similarity to personality traits. Specifically, in Study 2, children were shown triads consisting of a target character, a personality trait match, and a physical appearance trait match instead of the social category match. Once again, the personality trait match did not look like the target character, but had the same—verbally conveyed—personality trait as the target. The physical appearance trait match was also explicitly described as having the same physical trait as the target character (e.g., “they are both tall”), and in addition it had the relevant salient physical characteristic shared by the target. Thus, similar to Study 1, here too children faced an asymmetry regarding the format in which information about each test character was conveyed. However, whereas in Study 1 the alternative to personality trait was a psychologically relevant verbally and visually conveyed dimension (i.e., social category), in Study 2 the alternative was a psychologically not-so-relevant verbally and visually conveyed dimension (i.e., physical appearance trait). In other words, the only difference between Studies 1 and 2 had to do with the psychological relevance of the alternative dimension to personality trait.

If in Study 1, children’s tendency to infer novel properties by social category membership derived solely from the fact that they had received both verbal and visual cues regarding similarity on this category type, then children in Study 2 should show a preference for inferring based on physical appearance trait information. Alternatively, if children’s responses in Study 1 reflect the conceptual relevance of social category to children’s inferences, then children in Study 2 should not show a tendency to draw inferences based on physical appearance traits.

Method

Participants. Twenty-four adults (M = 22 years, range = 20–27 years) and 24 children (M = 5 years 6 months, range = 4 years 3 months to 6 years 4 months) participated in this study. None of the participants in Study 2 had taken part in Study 1. There were 20 women and 4 men among the adults, and 16 girls and 8 boys among the children. All adults were university students who either volunteered or received course credit for participating. Children were recruited from kindergartens in the cities surrounding the university. All children came from middle-class, secular, Jewish families. Only children with signed parental permission participated in the study.

Design. Age group (children vs. adults) was the only between-subjects factor. The study consisted of eight trials, each presenting children with the possibility to make an inference based either on the physical appearance trait or on the personality trait of the target character.

Materials. In each trial, children were shown a triad of line-drawn lightly colored human characters: a target character and two test characters. One of the test characters was presented to children as having the same physical appearance trait as the target but as having a different personality trait (appearance trait match). This character also resembled the target physically in terms of appearance trait markers (e.g., the “tall” character was clearly taller than the “short” character, a fact accentuated in the procedure
by aligning the characters). The other test character was presented to children as having the same personality trait as the target but as having a different physical appearance trait (personality trait match). This character did not resemble the target physically (see Figure 2 for an example).

Procedure. The procedure in this study was exactly the same as in Study 1. The only difference was that instead of social category information, children were told physical appearance information about one of the test characters. The four physical appearance traits used were selected because of their presumed familiarity to Israeli kindergarteners. They were “weight” (thin/fat), “height” (tall/short), “skin shade” (tanned/pale), and “hair color” (red hair/brunette). The four personality traits used, as well as the novel properties attributed to the characters, were the same as in Study 1.

Participants’ responses were scored as “1” if they inferred that the target character had the same property as the appearance trait match, and as a “0” if they inferred that the target character had the same property as the personality trait match. Thus, scores could range from “8” (all inferences according to physical appearance trait) to “0” (all inferences according to personality trait).

Results and Discussion

The main dependent measure used in the analyses was the number of trials, out of eight, in which participants answered that the target character had the same property as the appearance trait match.

In order to compare children and adults’ inferential patterns, we first conducted an ANOVA with age group and gender as between-subjects factors. This analysis revealed a significant effect of age group, with children making more physical appearance-based inferences than adults, $F(1, 44) = 23.81$, $p < .001$, $\eta^2 = .351$ (see Table 1). Neither the main effect of gender nor the interaction between gender and age group was significant. One-sample $t$ tests were conducted separately for children and adults in order to compare their pattern of responses to chance (chance = 4). These analyses revealed that children’s frequency of appearance-based inferences was not different than expected by chance ($p > .7$), whereas adults’ frequency was significantly smaller than

Figure 2. Example (in gray scale) of a “tall-short” stimulus set used in Study 2.
expected by chance, $t(23) = -8.33$, $p < .001$. Thus, adults systematically drew inferences based on personality traits.

Next, using the same classificatory system as the one described in Study 1, we classified participants as “personality trait reasoners” if they drew inferences based on personality traits on six or more of the trials, as “appearance trait reasoners” if they drew inferences based on physical appearance traits on six or more of the trials, and as “inconsistent reasoners” if they manifested a different distribution. Table 3 presents this distribution. Consistent with the findings from the ANOVA, there were more personality trait reasoners among adults than among children, and more inconsistent reasoners among children than among adults, Fisher’s exact test $p < .001$.

We next conducted separate $t$ tests on the frequency with which participants inferred based on each appearance or personality trait. Recall that for all participants, every physical appearance and personality trait appeared in two trials out of the eight. These analyses revealed that adults made appearance-based inferences significantly less often than expected by chance (chance = 1) on all physical appearance traits. In turn, the frequency of children’s appearance-based inferences was not significantly different from chance on all physical appearance traits, except hair color. Interestingly, for hair color children performed like adults, making appearance-based inferences less often than expected by chance ($M = 0.6$, $SD = 0.8$), $t(23) = -2.23$, $p < .05$. The complementary analyses on the inductive potential of each personality trait revealed that adults selected all personality traits significantly more often than expected by chance, and that children selected all personality traits no differently than expected by chance. Overall, these findings indicate that there were no substantial differences in the inductive power of either the various physical appearance or personality traits.

Last but not least, we compared the extent to which participants made personality trait-based inferences in Studies 1 and 2. This comparison addressed the power of social category information as opposed to physical appearance trait information, in swaying children and adults away from drawing inferences based on personality traits. Recall that both studies used the same personality traits and novel properties; thus the only difference was in whether personality traits were contrasted with social categories or with physical appearance traits.

An ANOVA with age and study as between-subjects factors revealed a significant effect of age, with adults making more personality trait-based inferences than children, $F(1, 92) = 49.41$, $p < .001$, $\eta^2 = .349$. More importantly, there was also an effect of study, with participants making more personality trait-based inferences in Study 2 than in Study 1, $F(1, 92) = 13.40$, $p < .001$, $\eta^2 = .127$. The interaction between the factors was not significant. These findings indicate that when inferring psychological properties, both children and adults were more likely to abandon personality traits if the alternative was social categories than if the alternative was physical appearance traits.

In general, the findings reinforce the theoretical position defended strongly by Hirschfeld (1996) that physical distinctiveness might not suffice for people to essentialize human kinds.

### Study 3

In Study 1, we found that visually and verbally conveyed social category membership won out against personality traits in determining children’s inferences. Study 2 showed that visually and verbally conveyed physical appearance traits did not beat personality traits in determining children’s inferences. We interpreted these findings as suggesting that children view social category as a conceptually deeper type of category for inferring people’s psychological properties than both personality and physical appearance traits. The goal of Study 3 was to evaluate the role of labels versus physical appearance in determining this supremacy of social categories for children. This evaluation is of great theoretical significance. As noted in the Introduction section, while Hirschfeld (1996), and especially Sperber (1996), emphasize the role of language as an essentializing factor, Gil-White (2001), and even more so Atran (1990), underline the physical distinctiveness component. This evaluation was also important to better understand the comparative power of social categories over both personality and physical appearance traits.

Regarding the comparison between social categories and physical appearance traits, Study 3 aimed to...
to evaluate the extent to which the familiar labels—as opposed to the physical appearance of the different types of categories—determine children’s inferences. It was possible that the difference between social categories and physical appearance traits found when comparing Studies 1 and 2 did not derive from conceptual differences between these categories but rather from differences in the degree of physical similarity of the characters presented to children. That is, perhaps the social category matches of Study 1 were seen by children as being more physically similar to the target character than were the appearance trait matches of Study 2. The difference in children’s responses in the two studies would thus simply reflect this physical similarity difference rather than a conceptual one.

As for the comparison between social categories and personality traits, the findings from Study 1 left open the possibility that social categories’ power derived in part from the fact that they were physically marked. The finding of Study 2 on physical appearance traits indirectly ruled out this possibility by showing that physical cues alone do not grant a category sufficient inductive power to beat personality traits. This finding notwithstanding, it was still possible that while not sufficient to explain why children viewed social categories as more inductively powerful than personality traits in Study 1, physical similarity might have contributed to this pattern.

In Study 3, we simply eliminated the variable of physical similarity by presenting children with schematic line-drawn faces of characters in which there were no physical markers of either social category or physical appearance trait. We presented children with pairs contrasting both social category and personality trait, and physical appearance and personality trait. In contrast to Studies 1 and 2, the information about social category and physical appearance trait was conveyed exclusively via familiar labels, just as was the information about personality traits.

This design helped clarify the comparison of social categories and physical appearance traits. If the difference in children’s responses to these categories in Studies 1 and 2 derived from systematic differences in their physical similarity, then in Study 3 there should be no difference in children’s tendency to infer novel properties based on social category compared with physical appearance trait. In turn, if the difference between Studies 1 and 2 reflects the privileged conceptual status of social categories relative to physical appearance traits, then the provision of familiar labels should be enough to trigger

This design also allowed us to compare directly and in a more balanced fashion the inductive potential of social categories versus personality traits. Specifically, if the supremacy of social categories over personality traits found in Study 1 derived in part by the former’s physical similarity component, then children in Study 3 should grant social categories less inductive power than did the children in Study 1. In turn, if the inductive supremacy of social categories over personality traits does not derive from the physical similarity aspect of the former, but rather from its conceptual relevance, then children in Study 3 should respond just as the children in Study 1 did. Given that these questions were primarily about children’s pattern of responses, Study 3 included only children as participants.

Method

Participants. Twenty-four children (M = 6 years 2 months, range = 5 years 5 months to 6 years 9 months) participated in this study. There were 13 girls and 11 boys. Children were recruited from kindergartens in the cities surrounding the university. All children came from middle-class, secular, Jewish families. Only children with signed parental permission participated in the study. None of the participants had taken part in the previous studies.

Design. “Category type” (social category vs. physical appearance trait) was manipulated within subjects. Specifically, on four of the eight total trials, each child was given the choice to make inferences based on either the social category membership or the personality trait of the target character. On the other four trials, each child was given the choice to make inferences based on either the physical appearance trait or the personality trait of the target character.

Materials. In each trial, children were shown a triad of schematic line-drawn faces of human characters. These characters had no distinctive visual markers indicative of their social category or their physical appearance trait (see Figure 3 for an example). One of the characters was the target; the other two were test characters. In the Social Category trials, the triads were structurally the same as those used in Study 1. Namely, one of the test characters was presented to children verbally as being of the same social category as the target but as having a different personality trait (social category match), whereas the other test character was presented to
children verbally as having the same personality trait as the target but as belonging to a different social category (personality trait match). In the Physical Appearance Trait trials, the triads were structurally the same as those used in Study 2. Namely, one of the test characters was presented to children verbally as having the same physical appearance trait as the target but as having a different personality trait (appearance trait match), whereas the other test character was presented to children verbally as having the same personality trait as the target but as having a different physical appearance trait (personality trait match).

Procedure. The procedure in this study was exactly the same as in Studies 1 and 2, including the counterbalancing measures. The only difference was that instead of eight trials of either social category or physical appearance trait information, all children had four trials of each. In each trial, children had a different social category or a different physical appearance trait contrasted with a personality trait. Thus each of the four social categories used in Study 1 and each of the physical appearance traits used in Study 2 appeared only once per participant in Study 3. The same four personality traits used previously appeared twice, once paired against a social category and once against a physical appearance trait. The novel properties were the same as those used in Studies 1 and 2.

Participants’ responses were scored as “1” if they inferred that the target character had the same property as the social category match or appearance trait match (depending on the trial type), and as “0” if they inferred that the target character had the same property as the personality trait match. Thus, scores ranged from “4” (all inferences based either on social category or on physical appearance trait, depending on the trial type) to “0” (all inferences based on personality trait).

Results and Discussion

In order to simplify the within-subjects comparison, the dependent measure used in this study was the number of trials in which participants answered that the target character had the same property as the personality trait match.

One of the goals of this study was to address whether or not familiar labels are enough to give rise to the differential inductive power of social categories and physical appearance traits. The findings from Studies 1 and 2 supported a unidirectional hypothesis that personality trait-based inferences would be more common in the physical appearance trials than in the social category trials (i.e., that children would be more likely to make inferences based on social categories than based on physical appearance traits). This hypothesis was tested, and confirmed, by a paired t test. Children were more likely to make personality trait-based inferences in the physical appearance trait trials (M = 2.08, SD = 0.97) than in the social category trials (M = 1.58, SD = 0.93), paired t(23) = 1.73, p < .05 (one tailed). In fact, analyses against chance (chance = 2) revealed that while the mean number of personality trait-based inferences children made in the physical appearance trials was not significantly different from chance, p > .6, in the social category trials it was
As in the previous studies, we next classified children as “personality trait reasoners” if they made personality trait-based inferences on three or four of the four trials of each type, as “nonpersonality trait reasoners” (i.e., either appearance trait or social category reasoners) if they made personality trait-based inferences on zero or one of the four trials of each type, and as “inconsistent reasoners” if they made personality trait-based inferences on two of the four trials of each type. As can be seen in Table 4, more children were “personality trait reasoners” in the physical appearance trials than in the social category trials, and the reverse was true in terms of “nonpersonality trait reasoners,” a difference found to be statistically significant by a sign test for related samples, p < .05. Overall, both the parametric and nonparametric analyses indicate that children were less likely to draw inferences based on personality traits when the foil was a social category than when it was a physical appearance trait, even though the only cue provided to children about category type was a familiar label.

As in the previous two studies, we again assessed whether there were differences among the individual social categories, physical appearance, and personality traits regarding their inductive potential. Recall that in this study, each social category or physical appearance trial appeared only once throughout the procedure. Thus, children could either draw an inference based on the nonpersonality trait category or not. Each personality trait, in turn, appeared twice. We found that the only social category in which children drew the inference consistently not based on personality traits was with ethnicity, χ²(1, N = 24) = 4.17, p < .05. Seventeen of the 24 children inferred that either two Jews or two Arabs would share the same novel psychological property, irrespective of their personality traits. Consistent with the findings from Study 1, the only other category in which the distribution approached significance was social status, χ²(1, N = 24) = 2.67, p = .10. Specifically, 16 of the 24 children inferred by social status and not by personality trait. For all the other social categories and physical appearance traits, selection frequencies did not approach significance, ps > .4. Analyses on the frequency of selection of each personality trait revealed that only for quietness (M = 0.7, SD = 0.6) the frequency differed significantly from chance, t(23) = −2.56, p < .05 (for all others, ps > .4).

Taken together, these findings suggest that differences in the familiar category labels seemed to account for the difference in the inductive power of social categories relative to physical appearance traits.

The other goal of Study 3 was to investigate whether or not familiar labels are enough to give rise to the differential inductive power of social categories compared with personality traits. First, the parametric analysis against chance reported above indicates that children were more likely to draw inferences based on social category than based on personality traits. A second analysis conducted to address this goal compared children’s responses in Study 3 with their responses in Study 1. This comparison allowed us to assess the incremental effect of physical markers to the inductive power of social categories.

Given that in Study 3 there were only four social category trials, whereas in Study 1 there were eight social category trials, we converted children’s responses in the two studies to the percentage of inferences based on personality traits, and then conducted an ANOVA comparing these percentages. In Study 1, 39% of children’s responses were of this kind; in the social category trials of Study 3, 40% were so—evidently a nonsignificant difference (p > .9). This finding suggests that the main factor driving the inductive power of social categories compared with personality traits was the familiar labels of the categories. The physical markers of social categories seemed to have no incremental effect on children’s inferences. (Interestingly, the same pattern was found when comparing the appearance trait trials of Study 3 with Study 2. In the appearance trait trials of Study 3, 52% of children’s responses were based on personality traits, while in Study 2, 51% were, p > .9.)

From a theoretical standpoint, these findings are consistent with Hirschfeld’s (1996) and Sperber’s (1996) emphasis on the role of language in pointing to children essentialized human kinds, and downplay the importance of physical markers in this

**Table 4**

<table>
<thead>
<tr>
<th>Trial Type</th>
<th>Personality trait reasoners</th>
<th>Inconsistent</th>
<th>Nonpersonality trait reasoners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social category trials</td>
<td>3</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>Appearance trait trials</td>
<td>9</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>

*In social category trials, nonpersonality trait reasoners refer to children who selected the social category match on three or four trials. In appearance trait-trials, nonpersonality trait-reasoners refer to children who selected the appearance trait match on three or four trials.
process, as noted by Gil-White (2001), and more forcefully by Atran (1990).

Study 4
The findings from the first three studies were taken to show that for 5-year-olds, social categories are a more conceptually powerful source of induction than are personality traits (and physical appearance traits). A question one might raise against this interpretation is whether the children at all considered personality traits. On the one hand, it has been argued that preschool aged children do not fully grasp the concept of personality traits (Alvarez, Ruble, & Bolger, 2001). On the other hand, Heyman and Gelman (2000) found that 5-year-olds could draw inferences based on shared personality traits. The goal of Study 4 was to address directly whether the low frequency of inferences based on personality traits made by the 5-year-olds in the first three studies derived from their incapacity to grasp personality traits, or from the relative weakness of this viable source of inference compared with the other sources.

To address this goal, Study 4 was designed as a complement to Study 3. In Study 4, we presented children with the exact same triads as the ones presented in Studies 1 and 2 (i.e., triads with relevant physical markers), only that we did not provide them with verbal information about the social category or the physical appearance traits of the characters. Instead, we provided children with information about the characters’ personality traits, and also told children that the two nonpersonality trait match characters shared proper names. The provision of proper names serves the methodological function of equating the labeling of the characters. And yet, the fact that two characters have the same proper name is not supposed to carry any implication as to them sharing psychological characteristics. In fact, 5-year-olds do not systematically draw inferences based on shared proper names (Heyman & Gelman, 2000).

If, as argued thus far, 5-year-olds can conceive of personality traits, but social categories are conceptually more powerful sources of induction, which derive their power from the familiar labels attached to them, then children in Study 4 should draw inferences based on personality traits. In turn, if personality traits are irrelevant pieces of information for 5-year-olds, then children in Study 4 should disregard that type of information, and draw inferences based on the physical similarity and shared proper names of the social category or physical appearance trait matches.

Method
Participants. Sixteen children (M = 5 years 9 months, range = 5 years 0 months to 6 years 4 months) participated in this study. There were 6 girls and 10 boys. Children were recruited from kindergartens in the cities surrounding the university. All children came from middle-class, secular, Jewish families. Only children with signed parental permission participated in the study. None of the participants had taken part in the previous studies.

Design. As in Study 3, “category type” (social category vs. physical appearance trait) was manipulated within subjects.

Materials. Four of the triads from Study 1 and four of the triads from Study 2 were used in this study (see examples in Figures 1 and 2). To avoid any familiarity effect, 16 non-Hebrew proper names were used (e.g., “Susan,” “Peter”).

Procedure. The procedure in this study was similar to that of Study 3, with the same within-subjects balancing of four social category and four physical appearance trait triads. The only difference was that in Study 4 the experimenter told children about the personality traits of the characters, their proper names, but did not mention their social category membership or their physical appearance trait. In each triad, the target character received the same personality trait label as the physically dissimilar test character (the personality trait match), and the same proper name as the physically similar test character (either the social category match or the appearance trait match, depending on the trial type). Scoring was done as in Study 3.

Results and Discussion
The main goal of this study was to address whether or not children could draw inferences based on personality traits. The dependent measure used was the number of trials in which participants answered that the target character had the same property as the personality trait match. We found that children were more likely to draw inferences based on personality traits (M = 5.56, SD = 1.55) than would be expected by chance (chance = 4), t(15) = 4.04, p < .005. Further analyses revealed that the mean number of personality trait-based inferences children made was significantly greater than chance (chance = 2) in both the physical appearance trials (M = 2.69, SD = 0.95), t(15) = 2.90, p < .05, and the social category trials (M = 2.87, SD = 0.96), t(15) = 3.66, p < .005. In fact, children were no more likely to make personality trait-based inferences in
the physical appearance trait trials than in the social category trials, $p > .5$. Notice that this latter null finding is consistent with the conclusion from Study 3 that physical similarity did not substantially differentiate between social category and physical appearance trait trials. Finally, analogously to the previous studies, we compared the frequency of selection of each personality trait to chance (chance = 1). We found that children selected both creativeness ($M = 1.6, SD = 0.5$), $t(15) = 5.00$, $p < .001$, and quietness ($M = 1.4$, $SD = 0.6$), $t(15) = 2.42$, $p < .05$, more than expected by chance.

The main conclusion from this study is that the low frequency of children's personality trait-based inferences in Studies 1-3 did not derive from their incapacity to rely on personality traits. Instead, it seems to have derived from the relative weak conceptual draw of personality traits compared with, primarily, social categories.

**General Discussion**

Preschool age children and adults draw inferences about people based on both their personality traits and social category membership (Heyman & Gelman, 2000; Hirschfeld, 1996). One of the main goals of the present studies was to investigate which of these two types of categorical information is a more powerful source of induction. Study 1 presented participants with the option to draw inferences about novel psychological properties of unfamiliar characters either by the characters' personality traits or by their social category membership. The results showed that while children tended to generalize these properties based on social category membership, adults did so based on personality traits.

One mundane explanation for why adults avoided drawing inferences based on social category membership is to do with social desirability or "political correctness." That is, perhaps adults believe it more appropriate to infer someone's preferences based on the person's personality traits than on his or her social category membership. Although this is indeed a plausible reason for the adults' pattern of responses, it does not explain children's tendency to infer based on social category membership. Plausibly, 5-year-olds are less sensitive to political correctness than adults. Nonetheless, this lack of sensitivity cannot explain why they have a preference for drawing inferences based on social category membership.

A possible explanation for this developmental pattern is that adults—especially Western adults—in addition to political correctness concerns, may have a belief that internal dispositions are more fundamentally related to people's preferences, desires, and behavioral tendencies. Children, in turn, may be more swayed by broad, culturally salient, categorical distinctions. It is only with the development of a more complete understanding of traits that children might come to realize their centrality in explaining individual people's psychological characteristics (Rhodes & Ruble, 1984; Yuill, 1992). This explanation is in line with Gil-White's (2001) account of the conceptualization of ethnic groups as essentialized species-like categories. In his view, in ethnic groups, like in animal categories but unlike in personality trait categories, members tend to marry within the group, and category membership is passed on from parents to offspring. These characteristics make ethnic groups—and perhaps other social categories—intuitive categories to the human mind.

The second main question driving the present studies was what factors contribute for social categories to have this intuitive and early-appearing conceptual power? To answer this question, we tackled two fundamental factors: the presence of physical markers and the presence of familiar labels.

Study 2 revealed that children's preference for social categories did not derive solely from the fact that social category information had physical correlates. In that study, children did not systematically draw inferences based on physical appearance traits, even though, like social categories, they were labeled and had physical markers. Study 3 showed that, in fact, physical markers did not seem to contribute at all to the inductive power of social categories. In that study, even though children had only the social category and personality trait labels to go by, their tendency to draw inferences based on social categories was statistically indistinguishable from the one revealed in Study 1, when in addition to labels children were also provided with social category physical markers. In general, then, labels, not appearances, carry more weight for social essentialism.

It is important to emphasize, however, that the conceptual work done by language is not solely a product of its formal characteristics. First, there is a difference between using a proper versus a common noun as labels. For instance, the attribution of the same social category noun to two Jewish girls (i.e., "Jewish") led children in Study 1 to infer that the girls had the same psychological characteristics despite having different personality traits. In contrast, the attribution of the same proper noun (i.e., "Susan") to the same two girls led children in Study 4 to overlook social category visual similarity and infer instead based on personality traits (see Heyman
Children were significantly more likely to draw inferences based on a common social category label than based on a common physical appearance trait label, and even within social categories certain labels were more powerful than others.

In regard to this latter point, in Study 1 we found that despite the overall developmental difference between children and adults regarding their preferred dimension for induction, children and adults seemed to agree on what were the most powerful social categories. Specifically, children’s tendency to draw inferences based on social category membership, in both Studies 1 and 3, was especially strong for the category of ethnicity (Jew/Arab), followed by social status (rich/poor). Adults’ tendency to draw inferences based on personality traits in Study 1 was especially weak for these same two categories. It is important to note that this type of consensus across studies and ages was not found with regard to either physical appearance or personality traits. For instance, whereas in Study 1 adults’ pull toward personality traits was especially strong with regard to activeness and niceness, children’s pull in Study 4 was especially strong with regard to creativeness and quietness. Thus, the consensus between children and adults regarding the inductive power of specific social categories does not seem to be arbitrary. By any count, the two most salient issues in the present-day Israeli agenda are security and economics. It is plausible, then, that ethnicity and social status owe their essentialist status to the nature of the cultural discourse.

Taken together, the present findings help delineate certain theoretical conclusions. First, the physical distinctiveness of a human group seems to play a negligible role in children’s tendency to essentialize the group. As Hirschfeld (1996) discussed, this type of finding contests bottom-up approaches to social essentialism, according to which children would arrive at essentialized human kinds by simply discovering them in their social environment. This finding also challenges claims about the importance of physical distinctiveness for the essentialization of human kinds (Gil-White, 2001), and arguments that the transfer of essentialism from the biological to the social domain is triggered by the physical distinctiveness of human groups (Atran, 1990).

Second, the present findings highlight the importance of labels in social essentialism. In a sense, labels seem to carry much, if not all, of the psychological meaning of social categories. What exactly is the role of language in the generation of social essentialism is debatable (Diesendruck, 2003). One possibility, advocated by Hirschfeld (1996), is that although the tendency to essentialize human kinds is innate, language or culture is needed to point out to children which specific human kinds to essentialize. Sperber (1996) argues that language is one of the main cultural tools by which essentialism itself gets applied to the domain of human kinds (see also Carey, 1995; Waxman, 1999). Somewhat inconsistent with this latter position, the present findings revealed that labeling per se is not enough. On the one hand, it is possible that some of the criteria noted by Gil-White (2001)—that is, within-kind marriage and inherited membership—might help explain why certain labeled kinds (e.g., social categories) are more essentialized than others (e.g., personality or physical appearance traits). On the other hand, in order to explain why particular human kinds that accord to Gil-White’s criteria get essentialized (e.g., ethnicity) whereas others do not (e.g., religiosity groups), we need additional factors. In this regard, Hirschfeld’s (1996) focus on the totality of cultural discourse is helpful, although future studies should attempt to break down this notion further into experimentally operational and manipulable variables.

Although the studies addressed only the inductive potential of particular exemplars of social categories, a further issue for future investigations is whether the pattern revealed in the studies is indicative of broader differences in how these categories are conceptualized. In the literature on conceptual development, inductive potential is commonly seen as one of the “symptoms” of essentialist thinking (see Gelman, 2003, for a review). Other symptoms include a belief that categories are stable and resistant to change, have innate potential, and are absolute rather than graded. Studies show that various social categories (e.g., race, gender, ethnicity) in different cultures (e.g., U.S., Mongolia, India, Madagascar) have one or more of these essentialist symptoms (Astuti et al., 2004; Gil-White, 2001; Hirschfeld, 1996; Mahalingan, 2003; Taylor, 1996). The implication for the present findings is that although they are consistent with the notion that ethnicity, and perhaps social status, is essentialized by Israeli children, we need further convergent evidence to establish this more firmly.

To what extent ethnicity is essentialized by Israeli children, and what exactly is the role of language and culture in this process, are important questions not only for what the answers might tell us about conceptual development, but perhaps also for how they might help us improve the complex social relations existent in the particular cultural context investigated here.
References


