stock colony had fewer surviving sisters or if adoptees were killed in tests), for a total of up to 25 trials per colony (40 in the all-adoptee colony); 115 trials in all.

The results confirmed the prediction of the queen discriminator hypothesis (Fig. 1). All non-nestmate sisters attacked each other violently, regardless of the proportion of adopted sisters in the mixed colony (Fig. 2B). Stock colony workers found that their sisters from the all-adoptee colony, containing a heterospecific queen but no nonsister workers, were just as unrecognizable as a single sister reared in a colony of heterospecific workers. Control reintroductions of mixed colony sisters to one another, immediately after their attacks on stock colony sisters, always resulted in acceptance (Fig. 1). In ten out of ten coded (blind) tests, workers of unknown origin could be assigned correctly to stock or mixed colonies on the basis of their reaction to stock colony sisters (*11*).

In the extensively studied small colonies of the primitively social bee, *Lasioglossum zephyrum*, the discriminators of individual nestmates are learned separately. Discriminators are not shared or transferred among nestmates; an unfamiliar bee is accepted if its odor is similar to that of any one nestmate (*3, 12*). In the very large colonies of highly eusocial ants and bees, it is unlikely that workers could learn every individual nestmate's odors. Some blend of discriminators must form a homogeneous label acquired and learned by all, as is demonstrated by the indiscriminate mutual acceptance of members of interspecific mixed colonies. Either a shared odor gestalt or transferable queen discriminators would serve this function; our data support the latter model (*13*). We do not suggest that workers lack discriminators, but that these are dominated by those of the queen, possibly to be "uncovered" on her removal or death. As a kin recognition system this mechanism is somewhat error-prone, vulnerable to incursion by slave-making ants and myrmecophiles (*14*). Normally, however, any individual enclosing in a colony is an offspring of its queen and is correctly so labeled by the acquisition of her discriminators; and any other ant sharing this label is usually a relative. Our results might also explain the apparently reduced efficiency of recognition in polygynous ants, if learning multiple queens' discriminators gives workers broader tolerances (*5*).

NORMAN F. CARLIN
BERT HÖLLDOBLER
Museum of Comparative Zoology,
Harvard University,
Cambridge, Massachusetts 02139

References and Notes

1. W. D. Hamilton, *J. Theor. Biol.* **7**, 1 (1964); E. O. Wilson, *Sociobiology* (Harvard Univ. Press, Cambridge, Mass., 1975).
2. R. Boch and R. A. Morse, *Ann. Entomol. Soc. Am.* **75**, 654 (1982); M. D. Breed, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2635 (1981); W. M. Getz, *J. Theor. Biol.* **99**, 585 (1982); C. P. Haskins and E. Haskins, *Psyche*, in press; A. Mintzer, *Behav. Ecol. Sociobiol.* **10**, 165 (1982); N. Ross and G. Gamboa, *ibid.* **9**, 163 (1981).
3. G. R. Buckle and L. Greenberg, *Anim. Behav.* **29**, 802 (1981).
4. R. Crozier and M. Dix, *Behav. Ecol. Sociobiol.* **4**, 217 (1979).
5. B. Hölldobler and C. D. Michener, in *Evolution of Social Behavior: Hypotheses and Empirical Tests*, H. Markl, Ed. (Verlag Chemie, Weinheim, 1980), pp. 35–58.
6. Colonies of *Camponotus americanus*, *C. ferrugineus*, *C. noveboracensis*, *C. pennsylvanicus*, and *C.* (*Mytmentoma*) *nearcticus* were collected in eastern Massachusetts. (Voucher specimens deposited in the Museum of Comparative Zoology, Harvard University.) Queens were housed in new test tubes (2.2 cm inner diameter, 15 cm long) containing water trapped at the bottom behind a cotton plug, placed in thoroughly washed plastic boxes (19.5 by 14 by 7.5 cm); mature colonies were given several test tubes in a larger arena. All were fed synthetic ant diet [A. Bhatkar and W. H. Whitcomb, *Fla. Entomol.* **53**, 229 (1970)], honey-water, chopped cockroaches, and *Tenebrio* larvae.
7. Kept at 30°C, these ants remain active through winter, but no new adults are produced. Evidence suggests that colony odors are learned or acquired (or both) by callow adults shortly after eclosion (*3*). Our colonies were tested 4 months after the summer-produced workers had passed this period.
8. A worker in a colony containing n_c conspecifics (including itself) and n_h heterospecifics may interact with $n_c - 1$ of the former and all n_h of the latter. If there is no preference for interacting with either, then for a total of A acts observed in a given species, the expected number of intraspecific and interspecific interactions are $e_c = A(n_c - 1)/(n_c - 1 + n_h)$ and $e_h = A(n_h)/(n_c - 1 + n_h)$, respectively. The significance of deviation of observed intraspecific and interspecific acts, A_c and A_h, from e_c and e_h was calculated from a χ^2 statistic with one degree of freedom.
9. This "pseudocaste" phenomenon is one in which members of mixed colonies show species-specific behavior differences (unpublished data). Data on antennation, allogrooming, and regurgitation are pooled in Table 1, but their significance is the same analyzed separately. In one of the two *pennsylvanicus-americanus* colonies, both species preferred to antennate and solicit regurgitation from *americanus*, while in the other *americanus* was solicited but not antennated preferentially. Allogrooming was indiscriminate among and between species in both colonies.
10. Aggressive behavior patterns varied in different species. For example, at level 3, *C. americanus* sprayed formic acid, while *pennsylvanicus* and *ferrugineus* grappled and bit but rarely sprayed. *C.* (*Mytmentoma*) *nearcticus* workers were nonaggressive, always avoiding other ants in 9-cm glass dishes. When forced to encounter each other by being confined in glass vials (0.8 cm in diameter, 5.5 cm long), these ants did attack non-nestmates.
11. We also observed level 3 aggression in each of six tests between genetic *C. americanus* sisters adopted by different *C. pennsylvanicus* queens. This excludes the possibility that species-specific, rather than colony-specific, discriminators are derived from the queen, since the workers had conspecific queens.
12. P. Kukuk, M. Breed, A. Sobti, W. Bell, *Behav. Ecol. Sociobiol.* **2**, 319 (1977); L. Greenberg, *Science* **206**, 1095 (1979).
13. Environmental odors specific to the colony may also serve as recognition cues [C. R. Ribbands, H. Kalmus, H. L. Nixon, *Nature* (*London*) **170**, 438 (1952); A. R. Jutsum, T. S. Saunders, J. M. Cherrett, *Anim. Behav.* **27**, 839 (1979)]. However, when such odors are controlled for, as in our protocol, recognition is usually not eliminated.
14. For example, R. K. Vander Meer and D. P. Wojcik, *Science* **218**, 806 (1982).
15. We thank D. Gladstein for formulating and testing the probability model, N. Franks and A. Robson for help collecting ants, and S. Bartz, P. Calabi, M. Reaka and E. O. Wilson for commenting on the manuscript. Supported by NSF grant BNS 80-02613 to B.H. and a Harvard University Richmond Fund grant to N.F.C.

22 April 1983; accepted 19 September 1983

Sex Differences in Mathematical Reasoning Ability: More Facts

Abstract. *Almost 40,000 selected seventh-grade students from the Middle Atlantic region of the United States took the College Board Scholastic Aptitude Test as part of the Johns Hopkins regional talent search in 1980, 1981, and 1982. A separate nationwide talent search was conducted in which any student under age 13 who was willing to take the test was eligible. The results obtained by both procedures establish that by age 13 a large sex difference in mathematical reasoning ability exists and that it is especially pronounced at the high end of the distribution: among students who scored* $\geq$ *700, boys outnumbered girls 13 to 1. Some hypothesized explanations of such differences were not supported by the data.*

In 1980 we reported large sex differences in mean scores on a test of mathematical reasoning ability for 9927 mathematically talented seventh and eighth graders who entered the Johns Hopkins regional talent search from 1972 through 1979 (*1, 2*). One prediction from those results was that there would be a preponderance of males at the high end of the distribution of mathematical reasoning ability. In this report we investigate sex differences at the highest levels of that ability. New groups of students under age 13 with exceptional mathematical aptitude were identified by means of two separate procedures. In the first, the Johns Hopkins regional talent searches in 1980, 1981, and 1982 (*3*), 39,820 seventh graders from the Middle Atlantic region of the United States who were selected for high intellectual ability were given the College Board Scholastic Aptitude Test (SAT). In the second, a nationwide talent search was conducted for which any student under 13 years of age who was willing to take the SAT was eligible. The results of both procedures substantiated our prediction that before age 13 far more males than females would score extremely high on SAT–M, the mathematical part of SAT.

The test items of SAT–M require numerical judgment, relational thinking, or insightful and logical reasoning. This test is designed to measure the developed mathematical reasoning ability of 11th and 12th graders (*4*). Most students in our study were in the middle of the seventh grade. Few had had formal opportunities to study algebra and beyond (*5, 6*). Our rationale is that most of these students were unfamiliar with mathematics from algebra onward, and that most who scored high did so because of extraordinary reasoning ability (*7*).

In 1980, 1981, and 1982, as in the earlier study (*1*), participants in the Johns Hopkins talent search were seventh graders, or boys and girls of typical seventh-grade age in a higher grade, in the Middle Atlantic area. Before 1980, applicants had been required to be in the top 3 percent nationally on the mathematics section of any standardized achievement test. Beginning in 1980, students in the top 3 percent in verbal or overall intellectual ability were also eligible. During that and the next 2 years 19,883 boys and 19,937 girls applied and were tested. Even though this sample was more general and had equal representation by sex, the mean sex difference on SAT–M remained constant at 30 points favoring males (males' $\overline{X} = 416$, S.D. = 87; females' $\overline{X} = 386$, S.D. = 74; $t = 37$; $P < 0.001$). No important difference in verbal ability as measured by SAT–V was found (males' $\overline{X} = 367$, females' $\overline{X} = 365$).

The major point, however, is not the mean difference in SAT–M scores but the ratios of boys to girls among the high scorers (Table 1). The ratio of boys to girls scoring above the mean of talent-search males was 1.5:1. The ratio among those who scored ≥ 500 (493 was the mean of 1981–82 college-bound 12th-grade males) was 2.1:1. Among those who scored ≥ 600 (600 was the 79th percentile of the 12th-grade males) the ratio was 4.1:1. These ratios are similar to those previously reported (*1*) but are derived from a broader and much larger data base.

Scoring 700 or more on the SAT–M before age 13 is rare. We estimate that students who reach this criterion (the 95th percentile of college-bound 12th-grade males) before their 13th birthday represent the top one in 10,000 of their age group. It was because of their rarity that the nationwide talent search was created in November 1980 in order to locate such students who were born after 1967 and facilitate their education (*8*). In that talent search applicants could take the SAT at any time and place at which it was administered by the Educational Testing Service or through one of five regional talent searches that cover the United States (*9*). Extensive nationwide efforts were made to inform school personnel and parents about our search. The new procedure (unrestricted by geography or previous ability) was successful in obtaining a large national sample of this exceedingly rare population. As of September 1983, the number of such boys identified was 260 and the number of girls 20, a ratio of 13.0:1 (*10*). This ratio is remarkable in view of the fact that the available evidence suggests there was essentially equal participation of boys and girls in the talent searches.

Table 1. Number of high scorers on SAT–M among selected seventh graders—19,883 boys and 19,937 girls—tested in the Johns Hopkins regional talent search in 1980, 1981, and 1982, and of scorers of ≥ 700 prior to age 13 in the national search (*9*).

Score	Number	Percent	Ratio of boys to girls
Johns Hopkins regional search			
420 or more*			
Boys	9119	45.9	1.5:1
Girls	6220	31.2	
500 or more			
Boys	3618	18.2	2.1:1
Girls	1707	8.6	
600 or more			
Boys	648	3.3	4.1:1
Girls	158	0.8	
National search			
In Johns Hopkins talent search region			
700 or more			
Boys	113	†	12.6:1
Girls	9	†	
Outside Johns Hopkins talent search region			
700 or more			
Boys	147	†	13.4:1
Girls	11	†	

*Mean score of the boys was 416. The highest possible score is 800. †Total number tested is unknown (*9*).

The total number of students tested in the Johns Hopkins regional annual talent searches and reported so far is 49,747 (9,927 in the initial study and 39,820 in the present study). Preliminary reports from the 1983 talent search based on some 15,000 cases yield essentially identical results. In the ten Middle Atlantic regional talent searches from 1972 through 1983 we have therefore tested about 65,000 students. It is abundantly clear that far more boys than girls (chiefly 12-year-olds) scored in the highest ranges on SAT–M, even though girls were matched with boys by intellectual ability, age, grade, and voluntary participation. In the original study (*1*) students were required to meet a qualifying mathematics criterion. Since we observed the same sex difference then as now, the current results cannot be explained solely on the grounds that the girls may have qualified by the verbal criterion. Moreover, if that were the case, we should expect the girls to have scored higher than the boys on SAT–V. They did not.

Several "environmental" hypotheses have been proposed to account for sex differences in mathematical ability. Fox *et al.* and Meece *et al.* (*11*) have found support for a social-reinforcement hypothesis which, in essence, states that sex-related differences in mathematical achievement are due to differences in social conditioning and expectations for boys and girls. The validity of this hypothesis has been evaluated for the population we studied earlier (*1*) and for a subsample of the students in this study. Substantial differences between boys' and girls' attitudes or backgrounds were not found (*5, 6, 12*). Admittedly, some of the measures used were broadly defined and may not have been able to detect subtle social influences that affect a child from birth. But it is not obvious how social conditioning could affect mathematical reasoning ability so adversely and significantly, yet have little detectable effect on stated interest in mathematics, the taking of mathematics courses during the high school years before the SAT's are normally taken, and mathematics-course grades (*5, 6*).

An alternative hypothesis, that sex differences in mathematical reasoning ability arise mainly from differential course-taking (*13*), was also not validated, either by the data in our 1980 study (*1*) or by the data in the present study. In both studies the boys and girls were shown to have had similar formal training in mathematics (*5, 6*).

It is also of interest that sex differences in mean SAT–M scores observed in our early talent searches became only slightly larger during high school. In the selected subsample of participants studied, males improved their scores an average of 10 points more than females (the mean difference went from 40 to 50 points). They also increased their scores on the SAT–V by at least 10 points more than females (*6*). Previously, other researchers have postulated that profound differences in socialization during adolescence caused the well-documented sex differences in 11th- and 12th-grade SAT–M scores (*11*), but that idea is not supported in our data. For socialization to account for our results, it would seem necessary to postulate (ad hoc) that chiefly early socialization pressures significantly influence the sex difference in SAT–M scores—that is, that the intensive social pressures during adolescence have little such effect.

It is important to emphasize that we are dealing with intellectually highly able students and that these findings may not generalize to average students. Moreover, these results are of course not generalizable to particular individuals. Finally, it should be noted that the boys' SAT–M scores had a larger variance than the girls'. This is obviously related to the fact that more mathematically talented boys than girls were found (*14*). Nonetheless, the environmental hypotheses outlined above attempt to explain mean differences, not differences in variability. Thus, even if one concludes that our findings result primarily from greater male variability, one must still explain why.

Our principal conclusion is that males dominate the highest ranges of mathematical reasoning ability before they enter adolescence. Reasons for this sex difference are unclear (*15*).

CAMILLA PERSSON BENBOW
JULIAN C. STANLEY
Study of Mathematically Precocious Youth, Johns Hopkins University, Baltimore, Maryland 21218

References and Notes

1. C. Benbow and J. Stanley, *Science* **210**, 1262 (1980).
2. Also see letters by C. Tomizuka and S. Tobias; E. Stage and R. Karplus; S. Chipman; E. Egelman *et al.*; D. Moran; E. Luchins and A. Luchins; A. Kelly; C. Benbow and J. Stanley, *ibid.* **212**, 114 (1981).
3. The Johns Hopkins Center for the Advancement of Academically Talented Youth (CTY) conducts talent searches during January in Delaware, the District of Columbia, Maryland, New Jersey (added in 1980), Pennsylvania, Virginia, and West Virginia. In 1983 coverage expanded northeast to include Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont.
4. T. Donlon and W. Angoff, in *The College Board Admissions Testing Program*, W. Angoff, Ed. (College Board, Princeton, N.J., 1971), pp. 24–25; S. Messick and A. Jungeblut, *Psychol. Bull.* **89**, 191 (1982).
5. C. Benbow and J. Stanley, *Gifted Child Q.* **26**, 82 (1982).
6. ———, *Am. Educ. Res. J.* **19**, 598 (1982).
7. We have found that among the top 10 percent of these students (who are eligible for our fast-paced summer programs in mathematics) a majority do not know even first-year algebra well.
8. J. Stanley, "Searches under way for youths *exceptionally* talented mathematically or verbally," *Roeper Rev.*, in press.
9. The regional talent searches are conducted by Johns Hopkins (begun in 1972), Duke (1981), Arizona State–Tempe (1981), Northwestern (1982), and the University of Denver (1982). Because there was no logical way to separate students who entered through the regional programs from those who entered through the national channel, results were combined. Most students fit into both categories but at different time points, since the SAT could be taken more than once to qualify or could be retaken in the regional talent search programs. The SAT is not administered by the Educational Testing Service between June and October or November of each year. Therefore, entrants who had passed their 13th birthday before taking the test were included if they scored 10 additional points for each excess month or a fraction of a month.
10. There is a remarkably high incidence of left-handedness or ambidexterity (20 percent), immune disorders (55 percent), and myopia (55 percent) in this group (manuscript in preparation).
11. L. Fox, D. Tobin, L. Brody, in *Sex-Related Differences in Cognitive Functioning*, M. Wittig and A. Petersen, Eds. (Academic Press, New York, 1979); J. Meece, J. Parsons, C. Kaczala, S. Goff, R. Futterman, *Psychol. Bull.* **91**, 324 (1982).
12. L. Fox, L. Brody, D. Tobin, *The Study of Social Processes that Inhibit or Enhance the Development of Competence and Interest in Mathematics Among Highly Able Young Women* (National Institute of Education, Washington, D.C., 1982); C. Benbow and J. Stanley, in *Women in Science*, M. Steinkamp and M. Maehr, Eds. (JAI Press, Greenwich, Conn., in press); L. Fox, C. Benbow, S. Perkins, in *Academic Precocity*, C. Benbow and J. Stanley, Eds. (Johns Hopkins Univ. Press, Baltimore, 1983).
13. For example, E. Fennema and J. Sherman, *Am. Educ. Res. J.* **14**, 51 (1977).
14. Why boys are generally more variable has been addressed by H. Eysenck and L. Kamin [*The Intelligence Controversy* (Wiley, New York, 1981)] and others.
15. For possible endogenous influences see, for example, R. Goy and B. McEwen, *Sexual Differentiation of the Brain* (MIT Press, Cambridge, Mass., 1980); J. Levy, *The Sciences* **21** (No. 3), 20 (1981); T. Bouchard and M. McGue, *Science* **212**, 1055 (1981); D. Hier and W. Crawley, Jr., *N. Engl. J. Med.* **306**, 1202 (1982); C. De Lacoste-Utamsing and R. Holloway, *Science* **216**, 1431 (1982); L. Harris, in *Asymmetrical Function of the Brain*, M. Kinsbourne, Ed. (Cambridge Univ. Press, London, 1978); M. McGee, *Psychol. Bull.* **86**, 889 (1979); S. Witelsen, *Science* **193**, 425 (1976); J. McGlone, *Behav. Brain Sci.* **3**, 215 (1980); D. McGuiness, *Hum. Nat.* **2** (No. 2), 82 (1979); R. Meisel and I. Ward, *Science* **213**, 239 (1981); F. Naftolin, *ibid.* **211**, 1263 (1981); A. Ehrhardt and H. Meyer-Bahlburg, *ibid.*, p. 1312; J. Inglis and J. Lawson, *ibid.* **212**, 693 (1981); M. Wittig and A. Petersen, Eds., *Sex-Related Differences in Cognitive Functioning* (Academic Press, New York, 1979).
16. We thank K. Alexander, L. Barnett, R. Benbow, R. Gordon, P. Hines, L. Minor, B. Persson, B. Polkes, D. Powers, B. Stanley, Z. Usiskin, and P. Zak. This study was supported by grants from the Spencer and Donner Foundations.

8 November 1982; accepted 17 May 1983

L-Tryptophan: A Common Denominator of Biochemical and Neurological Events of Acute Hepatic Porphyria?

Abstract. *Hepatic porphyrias are disorders of heme synthesis characterized by genetically determined lesions of one of the key enzymes of heme synthesis. In carriers of such lesions, several factors (drugs, environmental chemicals, or diet) precipitate acute and often fatal attacks of neurologic dysfunction, which are promptly relieved by intravenous infusion of heme. However, the mechanism of such heme-induced amelioration remains elusive. To probe this mechanism, the biochemical events triggered by acute hepatic heme deficiency were examined in an animal model of chemically induced porphyria. Acute hepatic heme depletion in porphyric rats was found to impair hepatic tryptophan pyrrolase activity which, in turn, elevated tryptophan and 5-hydroxytryptamine turnover in the brain. These alterations in porphyric rats were dramatically reversed by parenteral heme administration. These findings suggest that increased tryptophan and 5-hydroxytryptamine in the nervous system may be responsible for the neurologic dysfunctions observed in humans with acute attacks of hepatic porphyria.*

The three hepatic porphyrias—acute intermittent porphyria, hereditary coproporphyria, and variegate porphyria—are genetically transmitted disorders, each of which exhibits a defined defect of one of the enzymes essential for the formation of heme (*1*). The resulting heme deficiency in the liver removes the end-product (heme) repression of δ-aminolevulinic acid synthetase (ALAS), the first and rate-limiting enzyme of heme synthesis, leading to excessive formation and accumulation of heme precursors proximal to the particular enzyme block (*1*). Accordingly, depending on the site of the enzymatic defect, one may observe a pattern of heme precursor excretion that is characteristic for each of the three conditions (*1*).

In porphyric individuals, several factors, including exposure to a variety of common drugs or changes in hormonal status, diet, or fasting precipitate acute and often life-threatening attacks of neuropsychiatric dysfunction (*1*). Intravenous infusion of heme relieves or aborts these neurological manifestations (*1, 2*). However, the mechanism of this beneficial effect of heme is unknown. It is unlikely that administered heme directly substitutes for heme-deprived processes in the nervous system, because there is no evidence that infused heme enters the brain (*3*). The heme precursors δ-aminolevulinic acid (ALA) and porphobilinogen (PBG), which increase in concentration in the plasma of patients with acute forms of the hepatic porphyrias, produce neurotoxic effects in vitro (*4*), but there is no convincing evidence that they elicit neurological dysfunction when infused in large amounts in humans or animals (*5*). Moreover, neither ALA nor PBG appear to cross the blood-brain barrier to a significant extent (*5, 6*). The concentrations of these precursors in the brain or cerebrospinal fluid of porphyric patients are substantially below those required for demonstrable neurotoxicity in vitro (*7*). We therefore explored an alternative mechanism that might explain the heme-reversible neuropsychiatric dysfunction of acute heme deficiency in hepatic porphyrias.