Journal of Research in Personality 47 (2013) 953–957

Contents lists available at ScienceDirect

Journal of Research in Personality

journal homepage: www.elsevier.com/locate/jrp

Brief Report

At the interface of social cognition and psychometrics: Manipulating the sex of the reference class modulates sex differences in personality traits

Aaron W. Lukaszewski [a,*], James R. Roney [b], Michael E. Mills [a], Larry C. Bernard [a]

[a] 1 LMU Drive, Dept. of Psychology, Loyola Marymount University, United States
[b] Dept. of Psychological and Brain Sciences, University of California, Santa Barbara, United States

ARTICLE INFO

Article history:
Available online 5 September 2013

Keywords:
Gender differences
HEXACO
Personality
Personality assessment
Psychometrics
Sex differences
Social cognition

ABSTRACT

Psychometric surveys suggest that sex differences in personality are minimal. Herein, we argue that (a) the mind is likely biased toward assessing oneself relative to same-sex others, and (b) this bias may affect the measurement of sex differences in personality. In support of this, an experiment demonstrates modulation of sex differences on the HEXACO facets by manipulating the sex of the "reference class"—the group of people subjects compare themselves to when making self-assessments on survey items. Although patterns varied across traits, sex differences were relatively small in the "unspecified" and "same-sex" reference class conditions—but substantially larger in the "opposite-sex" condition. These findings point to a same-sex comparison bias that may impact the measurement of sex differences in personality.

1. Introduction

One of the major achievements of personality psychology has been the development of valid psychometric instruments for the measurement of individual difference constructs such as the Big Five (Costa, Terracciano, & McCrae, 2001) and HEXACO (Lee & Ashton, 2004) traits. The format for these personality surveys typically involves having subjects rate themselves on a variety of trait-exemplifying statements (e.g., "I avoid making small talk with people"), which are then combined into composite scores for each measured trait (e.g., sociability). Importantly, social comparison of oneself with others is one of the primary mechanisms through which self-evaluation occurs (Festinger, 1954; Gibbons & Buunk, 1999). As such, the act of assessing oneself on survey items may often involve comparative processes (although this could apply more to some items than others). For example, to the extent that social comparison is involved, rating a given trait-exemplifying statement as "very descriptive" of oneself would indicate that it is more descriptive of oneself than it is of others (Allen & Yen, 1979; Renick & Harter, 1989). Yet, despite the probable importance of social comparison in self-assessment, it is quite unclear to whom people actually compare themselves when completing survey items. It seems likely, for example, that young adults do not compare themselves with toddlers or people in a nursing home when they assess their relative standing on a personality trait. Given this, to which classes of people do subjects compare themselves?

In the current article, we illustrate how the answer to this basic social-cognitive question may have important implications for the ability of self-report surveys to measure sex differences in personality. First, we briefly review the prevailing consensus regarding sex differences in personality based on extant evidence from self-report surveys. Next, we argue that (a) the mind is likely biased toward assessing oneself in comparison with others of one's same sex, and if so (b) this same-sex comparison bias may affect the measurement of sex differences in personality. Finally, in support of these ideas, we describe an experiment that demonstrates modulation of sex differences on the HEXACO facet scales by systematically manipulating the sex of the *reference class*—that is, the group of people subjects compare themselves to when making self-assessments on survey items.

1.1. Does a same-sex comparison bias affect the measurement of sex differences in personality?

Based upon a large number of studies using validated psychometric surveys, researchers have concluded that sex differences in most aspects of personality are fairly minimal (Carothers & Reis, 2013; Costa et al., 2001; Feingold, 1994; Schmitt, Realo, Voracek, & Allik, 2008). Indeed, across the lower-order facets of the Big Five

* Corresponding author.
E-mail addresses: Aaron.Lukaszewski@lmu.edu, aalukas.1859@gmail.com (A.W. Lukaszewski), roney@psych.ucsb.edu (J.R. Roney), memills@gmail.com (M.E. Mills), Larry.Bernard@lmu.edu (L.C. Bernard).

0092-6566/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jrp.2013.08.010

traits, the average univariate effect size (d) for sex differences hovers in the small .25–.35 range (e.g., Costa et al., 2001). With the exception of the Emotionality factor, on which women score substantially higher than men, sex differences on the HEXACO facets are similarly small (Ashton & Lee, 2007; Lee & Ashton, 2004, 2006). Notably, this sex similarity holds even for traits that are differentially associated with one gender or the other. For instance, women and men score very similarly on the "assertiveness" facet of the NEO-PI-R (Costa et al., 2001) and the nearly identical "social boldness" facet of the HEXACO-PI-R (Lee & Ashton, 2004, 2006)—both of which capture agentic, socially competitive behavioral tendencies that are semantically associated with maleness and masculinity (e.g., Blair & Banaji, 1996).

This particular instance of apparent sex similarity is arguably surprising from multiple theoretical vantage points. From a social role theory perspective (e.g., Eagly & Wood, 1999), for example, the association of assertiveness and social boldness with traditionally male-biased roles (e.g., leadership; resource acquisition; and combat) implies that men should tend to be socialized such that they develop higher levels of these traits than women. Likewise, evolutionary theories hold that men have been subjected to a more intense ancestral history of direct intrasexual competition for status and resources than women, and should therefore have evolved to be bolder, riskier, and more interpersonally assertive on average, especially when the stakes of social competition are high (e.g., Archer, 2009; Buss and Schmitt, 1993; Campbell, 2002; Puts, 2010; Schmitt et al., 2008; Trivers, 1972).

Why, then, do women and men self-report similar levels of assertiveness and social boldness on personality surveys? One possible explanation, examined here, is that the mind is biased toward assessing oneself in comparison with others of one's same sex. Such a same-sex comparison bias is likely to exist for several reasons. First, because same-sex others have been one's most strategically relevant competitors for mates, resources, and social position over human evolutionary history (e.g., Daly & Wilson, 1983), it would have been functional ancestrally for assessments of one's own traits and abilities to be calibrated in relation to others of the same sex. Moreover, developmental research indicates that people typically spend more time in sex-segregated social groups and sex-typed occupational settings from early childhood onward (e.g., Serbin, Moller, Gulko, Powlishta, & Colburne, 1994; White & Brinkerhoff, 1981). As such, average members of each sex are likely to (a) form normative assessments of others' personality traits that are based differentially on observations of acts exhibited by same-sex others; and (b) derive knowledge of their own personalities from a set of behavioral acts they have exhibited more often in relation to same-sex others. In sum, these sorts of evolutionary, developmental, and social-cognitive considerations converge on the possibility that the mind may contain a bias toward same-sex comparison—which, if applicable to self-ratings on personality surveys, would naturally attenuate sex differences that might exist in actuality.

More specifically, from a psychometric standpoint, this argument suggests that personality surveys may not accurately capture sex differences because they leave unspecified the sex of the *reference class*—the group of people subjects compare themselves to when making self-assessments on survey items. The statistical logic underlying this hypothesis is straightforward: If men and women differ on a trait, but tend to assess themselves in relation to different, sex-biased reference classes by default, then sex differences that may actually exist will not be (entirely) evident in men's and women's group means based on self-ratings.

The purpose of the current study, therefore, was to preliminarily test the most straightforward empirical prediction generated from the above reasoning: that systematically manipulating the sex of the reference class would modulate the magnitude of sex differences on personality scales. To this end, we conducted a repeated measures experiment wherein subjects were asked to complete the HEXACO-PI-R facet scales (Lee & Ashton, 2004) in multiple different reference class conditions. In particular, subjects were asked to rate themselves on each survey item in comparison to (1) others in general (unspecified reference class condition); (2) others of the same sex (same-sex reference class condition); and (3) others of the opposite sex (opposite-sex reference class condition).

In general, the hypothesis of a same-sex comparison bias predicts that sex differences in the unspecified reference class condition will be most similar to those in the same-sex reference class condition (with average $|ds| < .35$), and that sex differences will be relatively large in the opposite-sex reference class condition. To the extent that this pattern obtains—either in general or for particular trait facets—this would provide evidence that people are biased towards same-sex comparison when responding to items on personality surveys.

2. Methods

2.1. Participants

Participants were 149 undergraduates from Loyola Marymount University (54 men; 95 women; *M* age = 19.09 years, *SD* = 1.68), who were issued partial course credit in exchange for participation.

2.2. Materials and procedure

Participants completed the experiment in a controlled laboratory environment in groups of 5–15. Upon arrival, each participant was seated at a private computer terminal, and instructed to carefully read the instructions before proceeding.

The repeated measures experiment simply involved taking the 100-item HEXACO PI-R (Lee & Ashton, 2004) three times: once in the "unspecified reference class" condition, once in the "same-sex reference class" condition, and once in the "opposite-sex reference class" condition. All participants first completed the unspecified condition and, thereafter, half completed the same-sex condition before the opposite-sex condition (and vice versa).

Instructions were identical across conditions except for the phrase that instructed participants to compare themselves to a particular reference class:

"On the following pages, you will find a series of statements about you. Please read each statement and determine the extent to which the statement is descriptive of you RELATIVE TO [OTHERS; SAME-SEX OTHERS; or OPPOSITE-SEX OTHERS]."

Items were presented in a random order, and participants responded to each item on a 5-point likert scale running from 1 (Much less descriptive of me than of [OTHERS; SAME-SEX OTHERS; or OPPOSITE-SEX OTHERS]) to 5 (Much more descriptive of me than of [OTHERS; SAME-SEX OTHERS; or OPPOSITE SEX-OTHERS]).

2.3. Data analyses

For each version of the survey, mean scores were computed for each HEXACO factor's four facets, plus the interstitial "altruism" scale. Table 1 presents these mean facet scores by sex and reference class condition, along with effect sizes for sex differences on each facet.

Analyses were conducted as follows: First, for each trait facet, we conducted a 2 (sex) $\times$ 3 (reference class) mixed ANOVA to determine whether sex difference magnitudes varied across reference class conditions. Second, we used Tukey *HSD* tests to evaluate effects of the reference class manipulation within each sex. Third,

Table 1
Mean sex differences on the HEXACO facet scales as a function of experimental reference class condition.

Factor	Facet	Sex	Reference class condition			Sex × reference class interaction
			Unspecified	Same-sex	Opposite-sex	
H	Sincerity	Men	3.28a	3.31a	3.45a	
		Women	3.17a	3.36b	3.12a	
		Sex difference (d)	.22	−.08	.47*	η^2 = .06*
	Fairness	Men	3.44a	3.48a	3.31a	
		Women	3.74a	3.51b	3.81a	
		Sex difference (d)	−.41	−.04	−.72*	η^2 = .10*
	Greed avoidance	Men	2.87a	3.03ab	3.16b	
		Women	2.80a	3.14b	3.20b	
		Sex difference (d)	.07	−.14	−.05	η^2 = .02
	Modesty	Men	3.21a	3.03b	3.18a	
		Women	3.36ab	3.40b	3.25a	
		Sex difference (d)	−.22	−.53*	−.10	η^2 = .06*
E	Fearfulness	Men	2.62a	2.66a	2.37b	
		Women	3.42a	2.91b	3.72c	
		Sex difference (d)	−1.02*	−.34	−1.88*	η^2 = .30*
	Anxiety	Men	3.47a	3.18b	3.08b	
		Women	3.85a	3.36b	3.61c	
		Sex difference (d)	−.54*	−.24	−.81*	η^2 = .04
	Dependence	Men	2.86ab	3.00a	2.68b	
		Women	3.49a	2.91b	3.65c	
		Sex difference (d)	−.80*	.02	−1.22*	η^2 = .25*
	Sentimentality	Men	3.07a	3.00a	2.69b	
		Women	3.71a	3.14b	3.82a	
		Sex difference (d)	−.84*	−.19	−1.58*	η^2 = .31*
X	Social self-esteem	Men	3.57a	3.29b	3.42a	
		Women	3.58a	3.42b	3.12c	
		Sex difference (d)	−.01	−.17	.42*	η^2 = .12*
	Social boldness	Men	3.00a	3.01a	3.16a	
		Women	3.02a	3.01a	2.89a	
		Sex difference (d)	.01	.00	.42*	η^2 = .05*
	Sociability	Men	3.34a	3.02b	2.93b	
		Women	3.43a	3.00c	3.25b	
		Sex difference (d)	−.12	.04	−.44*	η^2 = .07*
	Liveliness	Men	3.32a	3.08b	3.02b	
		Women	3.51a	3.28b	3.32b	
		Sex difference (d)	−.21	−.30	−.44*	η^2 = .00
A	Forgiveness	Men	2.79a	2.83a	3.07b	
		Women	2.48a	2.85b	2.66c	
		Sex difference (d)	.38*	−.03	.39*	η^2 = .09*
	Gentleness	Men	3.27a	3.24a	3.28a	
		Women	3.05a	3.18b	2.98a	
		Sex difference (d)	.31	.10	.43*	η^2 = .02*
	Flexibility	Men	2.85a	3.01ab	3.03b	
		Women	2.82a	2.98b	3.16c	
		Sex difference (d)	.05	.02	−.13	η^2 = .02
	Patience	Men	3.20a	3.20a	3.14a	
		Women	3.09a	3.12a	3.30b	
		Sex difference (d)	.13	.14	−.22	η^2 = .03
C	Organization	Men	3.30a	3.35a	2.91b	
		Women	3.71a	3.39b	3.96c	
		Sex difference (d)	−.45*	−.05	−1.28*	η^2 = .30*
	Diligence	Men	3.65a	3.49ab	3.43b	
		Women	3.79a	3.52b	3.57c	
		Sex difference (d)	−.20	−.05	−.20	η^2 = .01
	Perfection	Men	3.24a	3.21a	2.90b	
		Women	3.49a	3.21b	3.55a	
		Sex difference (d)	−.36	.00	−.93*	η^2 = .16*
	Prudence	Men	3.26a	3.18ab	3.07b	
		Women	3.40a	3.32a	3.46a	
		Sex difference (d)	−.20	−.20	−.53*	η^2 = .04
O	Aesthetic appreciation	Men	2.93a	2.95a	2.86a	

(*continued on next page*)

Table 1 (*continued*)

Factor	Facet	Sex	Reference class condition			Sex × reference class interaction
			Unspecified	Same-sex	Opposite-sex	
		Women	3.05a	2.98a	3.27b	
		Sex difference (d)	−.13	−.04	−.55*	η^2 = .06*
	Inquisitiveness	Men	2.94a	2.89a	3.04a	
		Women	2.56a	2.89b	2.74c	
		Sex difference (d)	.43*	.00	.43*	η^2 = .07*
	Creativity	Men	3.37a	3.21a	2.98b	
		Women	3.33a	3.10b	3.30a	
		Sex difference (d)	.05	.16	−.43*	η^2 = .12*
	Unconventionality	Men	3.35a	3.27a	3.14b	
		Women	3.12a	3.07a	3.07a	
		Sex difference (d)	.37	.36	.12	η^2 = .02
Interstitial	Altruism	Men	3.63a	3.36b	3.10c	
		Women	3.91a	3.55b	3.72c	
		Sex difference (d)	−.45*	−.30	−1.01*	η^2 = .11*

Note. For sex differences, positive values of Cohen's d effect sizes indicate that men are higher on average, while negative values indicate that women are higher on average. Effect sizes marked by asterisks (*) are statistically significant at $p < .01$ (as determined by independent samples *t*-tests for sex differences and 2 × 3 mixed ANOVAs for interactions). For within-sex comparisons across different reference class conditions, means that do not share a subscript within a row differ significantly at $p < .05$ (as determined by Tukey tests).

we used independent samples t-tests to test for sex differences on all trait facets in each condition.

3. Results and discussion

As can be seen in Table 1, there were significant sex x reference class interactions for 17 of the 25 facet scales, which indicate that the experimental reference class manipulation successfully modulated the magnitudes of sex differences in these traits. Moreover, these effects were substantial, explaining between 5% and 31% of the total variance in a given facet across sexes and conditions.

Consistent with our prediction, sex differences in the opposite-sex reference class condition were, on average, over four times larger (average $|d|$ = .63) than those in the same-sex condition (average $|d|$ = .14), and more than twice as large as those in the unspecified condition (average $|d|$ = .30). At a general level, then, the fact that sex difference magnitudes in the unspecified condition were much closer to those in the same-sex condition than those in the opposite-sex condition supports the hypothesis that people are biased toward assessing their own traits in relation to others of one's same sex.

However, the specific effects of the reference class manipulation on sex differences varied substantially across trait facets (Table 1). Most importantly for our hypothesis, on a number of the facets (e.g., sociability, social boldness, social self-esteem, and altruism), sex differences in the unspecified condition were very similar to those in the same-sex condition, but much larger in the opposite-sex condition. For these traits, then, the results provide clear evidence that people are biased toward same-sex comparison when the reference class is left unspecified.

Interestingly, the inverse pattern was found for a few other facets (forgiveness and inquisitiveness), such that sex differences in the unspecified condition most closely resembled those in the opposite-sex condition (with sex differences being either larger or smaller in the same-sex condition). As such, these results may indicate—counter to our general hypothesis—that people may actually be biased toward opposite-sex (rather than same-sex) comparison when rating items from certain traits.

For the remainder of facets (e.g., fairness and fearfulness), the magnitudes of sex differences in the unspecified condition were intermediate, somewhere in between those in the same-sex and opposite-sex conditions. As such, these findings may indicate that, for some traits, people do—as implicitly assumed by psychometricians—tend to compare themselves to others in general (i.e. from both sexes) by default.

Although we had no predictions in this regard, it is also noteworthy that the changes in sex differences across the conditions were in some cases driven by different patterns in men and women. To take one example, men reported nearly equal levels of fearfulness in the unspecified and same-sex conditions, but lower levels in the opposite-sex condition. Women, on the other hand, showed a much different pattern of fearfulness across conditions. Thus, these sorts of results not only have implications for the measurement of sex differences, but also point to potential sex differences in the content of self-assessment biases themselves.

3.1. Implications for personality assessment

The current study's suggestion of a same-sex comparison bias has a number of potential implications for the measurement of personality. Most provocatively, these findings are consistent with the notion that standard personality surveys may underestimate (or, in a minority of cases, overestimate) the magnitudes of sex differences on a variety of trait dimensions. Indeed, to take one example, if we were to solely rely on data from the unspecified reference class condition in order to test for a theoretically predicted sex difference in social boldness—as would typically be done in practice—we would conclude that average men and women do not differ on this trait. Crucially, though, results were nearly identical in the same-sex condition, but a substantial sex difference manifested when participants were asked to compare themselves to others of the opposite sex. Moreover, similar results were found for a variety of other HEXACO trait facets. Thus, although the results from the opposite-sex reference class condition cannot provide a precise estimate of the true sex difference magnitudes, our findings suggest that many sex differences could be larger and thus more important than recent theorists have claimed (e.g., Carothers & Reis, 2013).

However, another potential interpretation of the results is that, when asked to make sex-specific self-comparisons, participants compared themselves to stereotypes of men and women that are somewhat inaccurate. For example, if men are stereotypically (but not actually) more assertive than women, completing the same-sex and opposite-sex reference class conditions in direct

sequence may have primed subjects to assess their social boldness relative to inaccurate reference classes. If so, our results would illustrate how the implicit activation of gender stereotypes (e.g., Blair & Banaji, 1996) influences self-assessments on survey items. This too would be interesting from both a social-cognitive and a psychometric standpoint, but would not necessarily indicate that actual sex differences are being underestimated by standard personality surveys.

Testing between these possible (non-mutually exclusive) explanations of the current findings will be a critical task for future research that seeks to determine whether and how the putative same-sex comparison bias impacts personality measurement. For example, future research could examine sex differences obtained in different reference class conditions in relation to more objective measurements of trait-exemplifying behaviors. If self-reported sex differences were more similar to objective measures of behavior in the opposite-sex reference class condition than in the same-sex or unspecified conditions, this would provide compelling evidence that a same-sex comparison bias conceals sex differences in personality that exist in actuality.

Whichever of these explanations are correct, the current findings highlight a social-cognitive phenomenon that is seldom (if ever) explicitly acknowledged in psychometrics or personality psychology: that the act of self-assessing on survey items appears to involve implicit social comparison. This is especially important in light of the fact that standard psychometric surveys (including the HEXACO PI-R) do not instruct participants to compare themselves with others at all. This contrasts with the unspecified reference class condition in the current study, which instructed participants to assess themselves relative to "others". It is noteworthy, then, that the sex difference magnitudes obtained in our unspecified condition (average $|d| = .30$) are very close to those obtained in a larger university sample (average $|d| = .35$) using the standard instructions ($N = 1681$; Lee & Ashton, 2006). Moreover, an inspection of the trait-specific means and sex differences reinforces the apparent similarity of responses across these methods of instruction. Although it is possible this is coincidental, it seems more parsimonious to posit that this congruence reflects the fact that people self-assess relative to others on many items whether they are instructed to or not (as would be expected from the standpoint of social comparison theory; e.g., Festinger, 1954). If so, it would clearly be desirable for personality researchers to understand and gain control of these comparative processes.

4. Conclusion

In sum, the present findings provide preliminary evidence for a same-sex comparison bias that might partially conceal sex differences in self-reported personality traits. Although alternative explanations for the current results remain untested, our study may be considered an initial 'proof of concept' that highlights the potential importance of understanding which reference classes people compare themselves to when they complete survey items. Indeed, the same-sex comparison bias hypothesized herein is only one of many implicit reference class biases (e.g., those pertaining to age or status) that could in principle affect personality measurement. As such, it will be important for future research to further elucidate reference class effects of all kinds—which, if consequential for reliability and validity, could suggest the need for substantive changes to the design of psychometric surveys.

References

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole.

Archer, J. (2009). Does sexual selection explain sex differences in human aggression? *Behavioral and Brain Sciences, 32*, 249–266.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*, 150–166.

Blair, I. V., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology, 70*, 1142–1163.

Buss, D. M., & Schmitt, D. P. (1993). Sexual strategies theory: An evolutionary perspective on human mating. *Psychological Review, 100*, 204–232.

Campbell, A. (2002). *A mind of her own: The evolutionary psychology of women*. Oxford: Oxford University Press.

Carothers, B. J., & Reis, H. T. (2013). Men and women are from earth: Examining the latent structure of gender. *Journal of Personality and Social Psychology, 104*, 385–407.

Costa, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81*, 322–331.

Daly, M., & Wilson, M. (1983). *Sex, evolution, and behavior*. Monterey, CA: Brooks/Cole.

Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist, 54*, 408–423.

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin, 116*, 429–456.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117–140.

Gibbons, F. X., & Buunk, B. P. (1999). Individual differences in social comparison: Development of a scale of social comparison orientation. *Journal of Personality and Social Psychology, 76*, 129–142.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research, 39*, 329–358.

Lee, K., & Ashton, M. C. (2006). Further assessment of the HEXACO personality inventory: Two new facet scales and an observer report form. *Psychological Assessment, 18*, 182–191.

Puts, D. A. (2010). Beauty and the beast: Mechanisms of sexual selection in humans. *Evolution and Human Behavior, 31*, 157–175.

Renick, M. J., & Harter, S. (1989). Impact of social comparisons on the developing self-perceptions of learning-disabled students. *Journal of Educational Psychology, 81*, 631–638.

Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why can't a man be more like a woman? Sex differences in big five personality traits across 55 cultures. *Journal of Personality and Social Psychology, 94*, 168–182.

Serbin, L. A., Moller, L. C., Gulko, J., Powlishta, K. K., & Colburne, K. A. (1994). The emergence of gender segregation in toddler playgroups. *New Directions in Child Development, 65*, 7–17.

Trivers, R. L. (1972). The evolution of reciprocal altruism. *Quarterly Review of Biology, 46*, 35–57.

White, L. K., & Brinkerhoff, D. B. (1981). The sexual division of labor: Evidence from childhood. *Social Forces, 60*, 170–181.