Forthcoming in Evolutionary Psychological Sciences

Footbinding, hypergamy, and handicraft labor:

Evaluating the Labor Market explanation of footbinding

Ryan Nichols¹

¹Department of Philosophy, California State University, Fullerton

Author Note

Ryan Nichols https://orchid.org/0000-0001-9445-9821

Declarations: Funding: John Templeton Foundation award number 60704. Conflicts of interest/Competing interests: I have no known conflicts of interest to disclose. Ethics approval: Not applicable. Consent to participate: Not applicable. Consent for publication: Not applicable. Availability of data and material : Not applicable. Code availability: Not applicable. Authors' contributions: This is a single-authored submission. Not applicable.

Portions of this paper were presented in an invited lecture in the Behavior, Evolution, Culture series at University of California, Los Angeles, in April 2020.

Correspondence concerning this article should be addressed to Ryan Nichols, Department of Philosophy, CSU Fullerton, 800 N State College Blvd, Fullerton CA 92831 USA. Email: rnichols@fullerton.edu Blinded Manuscript (EXCLUDING AUTHORS CONTACT DETAILS) Click here to view linked References 1 RUNNING HEAD: Footbinding & Hypergamy

≛

Footbinding, hypergamy, and handicraft labor: Evaluating the Labor Market Hypothesis of footbinding

Abstract

Footbinding was a millennium-long practice in which ethnically Han Chinese families would alter the feet of young girls. The leading explanation of footbinding in interdisciplinary circles is the Labor Market Hypothesis. It posits that parents were motivated to footbind daughters to keep them spinning and weaving, thus making money for the family. Lacking direct means of testing their hypothesis, advocates of the Labor Market Hypothesis present allegedly disconfirming evidence against the major competing explanation of footbinding, an evolutionary social sciences hypothesis according to which footbinding is positively related to hypergamous marriage. This paper presents a methodological critique of this case against a hypergamy hypothesis for footbinding. This critique uncovers erroneous assumptions, methodological problems, ad hoc data modification, inaccurate operationalization and confounded testing of the hypergamy hypothesis, and concludes that this hypothesis ought not be rejected at this time.

Keywords: footbinding, hypergamy, China, evolutionary social sciences, labor market, interdisciplinary, Marxism

According to the Hypergamy Hypothesis for footbinding, parents used footbinding to increase the likelihood that their daughters would marry up at higher rates in open marriage markets more than would natural-footed girls. The Hypergamy Hypothesis fits within a larger evolutionary social sciences approach to footbinding. In a series of recent papers, evidence has been given against the Hypergamy Hypothesis (or 'HH') by advocates of an alternative theory, which we will call the Labor Market Hypothesis. According to the Labor Market Hypothesis (or 'LMH'), parents used footbinding to guarantee that daughters would stay indoors spinning and weaving so that parents could make money selling the handicrafts produced by their daughters. These two theories are not necessarily exclusive, but advocates of LMH have argued that HH is false and that, since HH is false, and since HH is LMH's only major competitor, the posterior probability of LMH rises.

The purpose of this paper is to provide readers a critical analysis of this case against HH published across several papers and books by the 'LMH team,' which consists primarily of coauthors Melissa J. Brown, Hill Gates, Laurel Bossen, and Damian Satterthwaite-Phillips. Due to opaque and tangled methods used by the LMH team to assess HH, this paper's critical assessment is complex. Table 1 provides a streamlined look at the case against HH and its flaws. Overall evidence presented by the LMH team does not decrease the probability of, let alone refute, the Hypergamy Hypothesis. Before entering into critical discussion, we first clarify the Labor Market Hypothesis (§1) and the Hypergamy Hypothesis (§2).

	Commitments of the Labor	D 11
Kesearch question	Market Hypothesis team	Problem of error
IJ ine пурегдату Untothosis (ШЦ) is		foothinding may instead be a post
disconfirmed does the	IIII is the most important	mitigation strategy for parents as they wait
aisconfirmed, does the	FIFI is the most important	findgation strategy for parents, as they wait
the Lahen Manhat	competitor to LMH. LMH	nor an auspicious marriage, rather than the
lhe Labor Markel	advocates reason that, if HH is	raison d'etre of footbinding. LIMH
Hypothesis (LMH) rise?	disconfirmed, then the posterior	advocates show no awareness of this
(§4)	probability of LMH rises.	alternative hypothesis and do not test it.
		HH states that footbinding is a means of
	The LMH team allegedly	enhancing one's mate value in contexts of
Can data from closed	disconfirm HH by using data	mate competition. It cannot be tested in a
marriage markets aptly	from closed marriage markets in	dataset absent both footbound and
test the relationship	which footbound women were	natural-footed women. Worse, the LMH
between footbinding and	not competing with natural-	team systematically removed data about
hypergamy? (§5)	footed women for grooms.	natural-footed women prior to testing.
Is wealth stratification		
among grooms and		Wealth stratification among grooms is a
grooms' families sufficient	The LMH team assumes that	necessary condition for testing HH but not a
to create conditions of	wealth stratification of families of	sufficient condition. Female mate
hypergamous marriage?	grooms is <i>sufficient</i> for proper	competition between footbound and
(§6)	testing of HH.	natural-footed women is also necessary.
		The LMH team converted their categorical
	The LMH team collected	data to ordinal data with no reason. The
	categorical data from surveys	LMH team refashioned their ordinal data
Does the LMH team	about women's memories of	into a 'marriage mobility index'. Authors
appropriately	wealth differences between their	adopt a convoluted and biased process for
operationalize HH for	natal and marriage families at time	operationalizing HH though a
testing? (§7)	of their marriage.	straightforward process is available.
0 10 /		A straightforward test involves tallying
		expected and observed values for
		hypergamous marriage for the 1139
		footbound and the 2575 natural-footed
		women in the sample, then calculating a
	Having modified the data in	loglikelihood test statistic to compare
	several ways the I MH team solit	ratios. The team's processes are
	3714 participants into many	convoluted ad hoc and violate common
	subcategories prior to testing	standards. These include not correcting for
	including 3 for hypergamy 2 for	familywise error and treating a region with
Does the I MH to and	footbinding categories 4 for	six footbound women as evidentially
abbrobriately test UU?	racion Authors than run numbers	acuivalent to a region with 004 foothourd
(68)	of statistical tests	women
(J0)	or stausucar tests.	WUIIICII.

Table 1. Problems with Brown.	Bossen.	Gates, et al.'s	case against the	Hypergamy	Hypothesis
	Dooten,	Sates, et al. s	ease against the	1 JPC Sunny	rypourcoio

What is the Labor Market Hypothesis for footbinding?

LMH explains footbinding in terms of microeconomic forces at play in the Han Chinese family and the self-interests of parents. Young girls spinning thread and weaving cloth inside the home made money for parents. Advocating a neo-Marxist position, LMH says footbinding owes its *origins* (Bossen & Gates 2017, 4, 186) and *maintenance* (see below) to the exploitation of daughters' handicraft labor by parents who extracted monetary value from daughters' work. Footbound girls spun cotton, hemp, and silk thread and yarn, and weaved cloth and hemp articles. Parents the world over have benefitted from labor of their children, but rarely did they find it necessary to physically hobble them, often permanently, or put them at grave risk of medical and health complications as in footbinding.¹ So why was footbinding necessary for parents' use of daughters' labor? Members of the LMH team clarify that "five-year-olds like to run and play, not sit in one place... By making it hurt to walk, footbinding served as labor control to keep girls at their handwork" (Brown 2016, 517), and add, "footbinding ensured that young daughters sat still and worked at various forms of tedious hand labor" (Bossen and Gates 2017, 10, 12).

Advocates of LMH have two types of evidence for their theory that it was parents' narrow monetary self-interest that initiated and maintained footbinding practice. The first is correlational evidence and the second, as mentioned, is evidence that HH is false. The correlational evidence is drawn from two sources: relationships between rates of footbinding and rates of handicraft

¹ In but one example of multiple sources of such complications, orthopedic surgeons Fang and Yu (1960) dissected bound feet and described them as akin to a "calcaneo-cavus foot of paralytic origin" (197) in which the inclination of forefoot and metatarsals were double that in a natural-footed woman.

production by daughters, and from correlations between the timing of the cessation of footbinding in China with the timing of the arrival of railroads, which brought cheap machined cloth. In terms of the first batch of correlational evidence, Bossen and Gates conduct a logistic regression on data aggregated about the correlation between handicraft work and footbinding. They conclude that the odds that footbound girls would do handicraft work were 2.14 times the odds that a girl who did not do handicraft work would be footbound (2017, 143). As to the second form of correlational evidence, a range of historical facts are used to justify the claim that trains brought cheap machined cotton and yarn to footbinding regions just before the market for handicrafts crashed. Rather than merely a correlation, this is alleged to have caused a sharp drop in parents' incentive for footbinding. In contrast to this purported explanation stands a well-known alternative according to which footbinding ceased because of the work of anti-footbinding or natural-foot societies (Drucker 1981).

Several facts come together to motivate a thorough analysis of the Labor Market Hypothesis and its case against the Hypergamy Hypothesis. First, correlational evidence for LMH is fragile and tenuous. The logistic regression noted above is the only inferential test statistic presented by Bossen and Gates in their recent book about footbinding. Other statistics in the book (there are many) are descriptive, raising questions about methodological rigor. Related, it is plausible to suppose that natural-footed girls had lower odds of doing handicraft labor than did footbound girls because the natural-footed girls were doing *ather forms of labor of greater value to the family* than spinning, like efficient agricultural work, which required healthy feet. The LMH team also does not show that the wellknown alternative explanation for footbinding's cessation is erroneous. Second, it appears the LMH team across their many publications neglects to statistically test an obvious competing hypothesis that daughters' handicraft labor was a cost-mitigation strategy used by parents to recoup some of the many losses due to material and opportunity costs associated with footbinding a daughter for years. Third, there are no publicly available datasets about footbinding. Since their data collection began in 1991, and after over a dozen publications about footbinding since, LMH team members have yet to release their data for supplemental testing despite requests (personal communication, 17 July 2017, 21 November, 2017, etc.). This appears to breech policies of funders of the authorship team's research, including the National Science Foundation (NSF 2020; cf. BCS#0613297, Melissa Brown, PI). These facts heighten the significance of a thorough review of the case against the Hypergamy Hypothesis given by Brown, Gates, Bossen, Satterthwaite-Phillips and others.

2 What is the Hypergamy Hypothesis, and what is its relation to an evolutionary sciences approach to footbinding?

The second major theory about the origins and maintenance of footbinding is represented broadly by an evolutionary sciences approach and narrowly by the Hypergamy Hypothesis. At least since M. Dickemann's groundbreaking papers on female claustration, female genital mutilation, and footbinding (1979, 1991), evolutionary social scientists generally understood footbinding to be produced in significant part through dynamics of mate competition, paternal certainty, and hypergamous marriage. Hypergamous marriage has been widespread, if uneven, due to evolved mate preferences in men for women who are youthful, fertile and beautiful, and mate preferences in women for men who have the status and wealth to provide for them and their offspring (Dickemann, 1979, 1997). Rates of hypergamy tend to increase in societies that are highly stratified, have low gender equity, and practice polygyny (Hartung, Dickemann, et al., 1982). This describes traditional China. In such locales, men who have multiple wives have the resources to outbid men of reduced means for the best young brides, and young women will often seek marriage as a second or third wife of a powerful man rather than a first wife of a poor man. This in part explains why sinologists' default assumption is that girls' feet were bound not due to a profit motive of parents but because doing so increased daughters' marriage prospects (e.g. Blake, 1994, 702). Hypergamous marriage plays a featured role within the set of evolutionary concerns relevant for explaining footbinding, but is only one of several. From a broad evolutionary perspective, highstatus polygynous men faced a unique set of mating challenges. Among these concerns, polygynous men, *ceteris paribus*, would find ensuring paternal certainty to be more demanding than would monogamously married men (see Buss & Shackelford, 1997). Whether polygynous or monogamous, paternal certainty is desired for the groom by both the groom himself and his parents. (It is parents who made marital decisions for children in historical China.) Since marriage functioned in accord with market-based principles of assortative mating, families of brides had incentive to inculcate in their daughters behaviors, virtues, and traits that reduce the perception of risk of paternal uncertainty for grooms and their parents. Display of these traits are hypothesized to increase daughters' mate value in open marriage markets. Footbinding probably arose from this state of affairs. It is likely that from mate preferences of this group of elites, footbinding was culturally transmitted to other socio-economic classes. (This represents the actual direction of its transmission in history according to a leading historian of footbinding; see Ko 2005, 113-115.)

Previous evolutionary social sciences explanations of footbinding leave a great deal to be desired. Current work is limited, relevant explanations are not unified, and this work appears uninformed by relevant facts about the historical practice of footbinding. These facts include elaborate ranking systems of the size and beauty of footbound feet, which appear to support the Hypergamy Hypothesis, and significant problems with general use of the term 'footbinding'.² One

² The coarse-grained term 'footbinding' is problematical. Recent historical work on Qing Dynasty footbinding practice (Gao, 2007) analyzes taxonomies of bindings and of feet from that period, originally proposed by Fang Xuan. This involved 'Five Styles' (五式), 'Eighteen Names' (十八名), and 'Nine Levels' (九品) (Gao 2007, 63-70). The 'Fake Level Lower-Lower' (赝品下下), the lowest

area in which the broad evolutionary sciences approach to footbinding has excelled, however, is in an understanding of game-theoretic mechanisms that sustained the maintenance of the practice over time, even in locales in which nearly all women were footbound and mate competition is reduced. Underscoring the dynamical element of an evolutionary approach, Gerald Mackie (1996) observed that parents have incentive not to be the first people in such a social ecology to refrain from footbinding their daughters for fear that they would be unmarriageable. In this context, evolutionary social scientists observe socio-functional features common between footbinding and female genital cutting, an evolutionarily stable marriage strategy for families of daughters (Howard & Gibson, 2017).

3 What is the relationship between the Labor Market and the Hypergamy Hypotheses?

The Labor Market Hypothesis and the evolutionary sciences approach are logically compatible in several respects, a fact that generates theoretical complexities yet to be appreciated. Both entail that footbinding reduces mobility, for example. LMH and the evolutionary sciences approach might implicitly appeal to parent-offspring conflict theory. Thankfully the two make

of the nine recognized levels of footbound foot, is described as having a high heel and being pointed but not thin. This foot would be appraised as spectacularly unappealing to matchmakers and grooms' families as compared to the 'God-level Upper-Upper' (神品上上) foot. Suppose hypotheses drawn from the evolutionary sciences, including HH, were to be tested with a pool of girls bound into 'Fake-level Lower-Lower' feet, or in a 'Passion Flower' style, as if such groups represent footbound women generally. These hypotheses would be wrongly falsified. This serves as only one example of historical complexities and one use of Chinese-language scholarship that ought to, but has yet to, inform evolutionary social scientific research on footbinding. Note that the LMH team reports no information about size or shape of their participants' feet at time of marriage. Footbinding & Hypergamy

opposite predictions in a pivotal area, marriage. Only if the evolutionary sciences approach is correct is it likely that (i) footbound girls would fetch larger brideprices than natural-footed girls, *ceteris paribus*; (ii) footbinding would be an arena of female mate competition as evidence, e.g., in beauty contests for footbound feet or ranking systems for footbound feet (see footnote 2); (iii) men would hypersexualize bound feet; (iv) feet would be frequently bound in ways causing more physical damage than needed to sequester a girl's handicraft labor; (v) bound feet would receive more care and attention just prior to marriage than at other times of a girl's life; and, of course, (vi) girls who are footbound would marry up at higher rates in open marriage markets, *ceteris paribus*, than natural footed girls.

The broad compatibility of the two theories means that exceptional attention is given to those entailments on which the two differ. It is for this reason the LMH team mounts an extensive rearguard defense of their theory by arguing at length against the Hypergamy Hypothesis, (vi) above. Advocates of LMH conclude across publications that footbinding did not correlate with hypergamy. Brown writes, "Subsequent analysis of whether footbinding improved a woman's chances of marrying to a wealthier family showed that the majority of women across ten provinces--whether footbound or not--married at the same economic level as their natal household" (2016, 515; see also Brown, et al., 2012, 1040). This fuels the reasoning that, since HH is its major competitor, if HH is disconfirmed, then the posterior probability of LMH rises.

4 If the Hypergamy Hypothesis is disconfirmed, does the posterior probability of the Labor Market Hypothesis rise?

In such an impoverished evidential environment, the LMH team's 'rearguard defense' plays an outsized role in the case for LMH. A debilitating problem with this reasoning is observed only once we consider the aforementioned cost-benefit analysis of footbinding. If footbound girls (and/or their natal families) *do not* benefit through a higher rate of hypergamous marriage *and* there are costs to footbinding, then footbinding is likely only to be maintained by convention. If footbound girls (and/or their natal families) *do* benefit through hypergamous marriages *and* there are costs to the practice, then footbinding will be maintained *so long as practices are enlisted to reduce those costs*. In this scenario, the costs of footbinding can be aptly minimized by putting footbound daughters to work while parents await an auspicious marriage for them. See Table 2.

Table 2. Handicraft labor as a *cost-minimization strategy* and not the *raison d'être* of footbinding

		Are there benefits to footbound girls who marry up?		
		Yes	No	
Are there costs	Yes	Conditionally expect	Do not expect	
to footbinding		footbinding, depending	footbinding	
daughters?		on magnitude of costs		
	No	Expect footbinding	Neutral model	

LMH considers the primary benefit of footbinding to be parents' accumulation of wealth through child labor while the daughter resides in the natal home. The evolutionary sciences approach considers the proximate benefits of footbinding the reduction of paternal uncertainty in grooms' families, better relative positions in open marriage markets for daughters, and correspondingly higher rates of hypergamous marriage than possessed by the natural-footed, and the ultimate benefit is greater fitness. From the evolutionary approach, income from labor of footbound girls may be an effective cost-mitigation strategy.

Footbinding is costly, especially for the poor mothers who composed the sample in LMH team data. Opportunity costs are paid by mothers and grandmothers who sacrifice more productive labor to clean and rebind daughters' feet every few days. A similar cost is incurred insofar as a bound daughter (depending on the severity of the binding) is unlikely to be able to perform chores requiring unimpeded mobility. Material costs were significant, including cloth bandages, bindings, antiseptic additives, and expenses paid to doctors and Traditional Chinese Medicine pharmacists to treat gangrene, infections, injuries from falls, and medical complications. In "Bound feet: How sexy were they," Hill Gates writes, "at least half of the little girls forced to endure it had to be beaten into submission by their elders" (2008, 58), which gestures at psychological distress in both parties. The residual emotional trauma experienced by mothers and grandmothers who caused extensive pain to their unwilling daughters must have been paid for in a currency and at a cost all its own. These costs are all multiplied across a decade of footbinding prior to marriage. Parents would be instrumentally irrational not to mitigate these costs. Therefore, *even if advocates of LMH used an unblemished dataset to prove the null hypothesis that footbinding is not positively correlated with hypergamy, this does not significantly raise the probability of LMH.* To justify such an inference, the LMH team must instead show that footbinding is not merely a cost-mitigation strategy. Its authors do not show awareness of this alternative let alone test it.

5 Can data from closed marriage markets aptly test the relationship between footbinding and hypergamy?

A marriage market is a social space in which those interested in finding a husband or wife, or their representatives, communicate to identify prospective partners, scout other contenders, selfassess mate value, detect trends, negotiate terms, and work with third-party brokers. To say a marriage market is *closed* is to note some or other restriction upon competition. In a *de jure* restriction, a marriage market in the Indian subcontinent might exclusively accept Brahmin caste members, in which case it is closed to Dalits. In a *de facto* restriction, due to transportation difficulties getting to a hilltop public square, the market held there is effectively closed to people lacking capable transport. Open markets are those that permit competition from relevant groups, where 'relevant' is defined in terms of the hypothesis at issue. The openness of a market is a matter of degree.

If footbinding saturated a population at a place and time, say where 95% of eligible girls were footbound, Chinese parents knew footbinding would not guarantee hypergamous marriages.

Where the population of marriageable women in a given locale was composed of the exclusively footbound, the market was *de facto* closed to natural footed women. For centuries in various pockets of China, footbinding persisted in countless *de facto* closed marriage markets. This is plausibly because parents believed footbinding daughters was a necessary condition for marrying off their daughters at all.

At this inferential juncture an evolutionary approach makes use of the game-theoretic analysis of what Mackie (1996) calls the "self-reinforcing" character of footbinding introduced above. Footbinding becomes *self-reinforcing* because parents feared catastrophic disutility were they to forego binding daughters. Parents believed that risks of an unmarriageable daughter, i.e. no grandchildren by her, were not worth the benefits of not footbinding. Chinese parents were burdened to balance disutility caused by footbinding against considerations about the fitness and life-satisfaction of their natural-footed daughters, daughters who would, as they saw it, never marry and never have children.

On the basis of the foregoing reasoning, any charitable version of HH must contain a crucial commonsensical qualification: *in open marriage markets* footbound girls will be more likely to marry hypergamously than natural-footed girls. Advocates of LMH launch their attack upon HH by omitting this condition. Worse, the LMH team made the befuddling decision to perform a *post hoc* modification of their data in a way relevant to the openness of marriage markets, as discussed in the following inductive argument.

Assumption. In their case against HH, Brown, Bossen, Gates, and co-authors report removing data from their survey results prior to testing. The modified dataset against which they tested HH, they write, only included data from "selected women with bound feet who were born in five-year birth cohorts *in which the majority of girls born in that cohort were subjected to binding*" (Brown et al., 2012, 1048; my italics). They describe this process as using "majority cohorts." Whether footbound or natural-footed, if a woman was born in a cohort wherein only 49% of girls were footbound, LMH team members systematically removed their data from the dataset used to test HH.

First premise. The LMH team conducted fieldwork in many counties across four 'regions' in China, Central, North, Southwest and Sichuan. At the county level, consider Ding County and its villages as an example to understand the structure of their data. They report that of 492 girls in Dongting Village's 1880s birth cohort, 99% were footbound; of 212 girls in Dongting Village's 1890s cohort, 94% were footbound; and of 21 girls in Qingfengdian-Pang Village's 1890s cohort, 100% were footbound. At the regional level, consider the North region and its nine field sites. Here the LMH team split their survey participants up into birth cohorts of five years each. (The length of "birth cohorts" varies without stated reason; see Bossen & Gates 2017, 178.) These nine sites yielded a total of 22 sample groups born in 1910-1914, 1915-1919 and 1920-1924. For twelve of the 22 cohorts in this region, that is, for over half of the cohorts tested, a full 100% of women surveyed were footbound as girls (Bossen & Gates, 2017, 179). In the Southwest region, birth cohorts from 1910-24, 1925-1929 and 1930-1934 yielded fifteen sample groups across five field sites. Here five of these fifteen groups had rates of footbinding over 90% (Bossen & Gates, 2017, 180). (Central and Sichuan regions are singled out for specific discussion below.) The LMH team say they tested HH by using survey data from women who were in birth cohorts in which 51% or more of the women were footbound. According to information about birth cohorts in appendices however, the survey data they actually used to test HH were from cohorts in which nearly all participants were footbound. Our first premise therefore reads: Cohorts included in data used to test HH were often saturated with footbound women so that the representation of natural-footed women in a given cohort was nil or nearly nil.

Second premise. Marriages among participants in the study were hyper-local. Bossen and Gates write, "Marital homes were generally within walking distance of natal homes, within a ten to fifteen-kilometer radius" (2017, 34). In her book Chinese women and rural development (2002) Bossen attests that

daughters married patrilocally but maintained regular contact with their natal families, and that the matchmaking process required many face-to-face meetings by the matchmaker.

Third premise. Girls who were in birth cohorts with X% of footbound women competed for grooms with other girls in the same birth cohort in villages <15km away, where ~X% of whom were also footbound. Put otherwise, if 95% of eligible girls in birth cohort 1 from village A are footbound, we can safely infer that roughly ~95% of eligible girls also in birth cohort 1 but who live in villages B-E, which surround village A at a distance of less than 15 kilometers, are also footbound. Thus, *women whose survey responses were included in the test of HH were almost exclusively footbound and these women were in competition with women from nearby villages who were also almost exclusively footbound*.

Fourth premise. Fourth is a premise and an intermediate conclusion implied by the first three premises. This is also of special importance for understanding the likelihood of footbound women marrying hypergamously. *In the modified dataset used by the LMH team to test HH, the vast majority of grooms' families must select local footbound brides*. Whether wealthiest, poorest, or in between, nearly all grooms' families must select a footbound girl as bride in the LMH team's modified dataset.

Fifth premise. As a result of 1-4, data used to test HH represented closed marriage markets. Women whose data were tested are exactly those not competing with natural-footed women for grooms. The rates at which footbound women in the tested sample married up cannot be meaningfully compared to the rate at which natural-footed women married up.

Conclusion. Brown, Bossen, Gates, and co-authors affirm the null hypothesis that there is no correlation between footbinding and hypergamy and conclude HH is false. However, to test HH, one must test this relationship in an open marriage market in which footbound daughters are competing against natural-footed daughters for grooms. *Due to inappropriate assumptions, unwarranted post hoc data modification, and sampling error, the LMH team's case against HH is implausible.*

6 Is wealth stratification among grooms and grooms' families sufficient to create conditions of hypergamous marriage?

One might object that the Hypergamy Hypothesis is testable with the modified data used by the LMH team on the condition that there is sufficient wealth and status stratification among families of grooms. This is to object to premise four above.

To understand why this objection lacks merit, distinguish *necessary* from *sufficient* conditions for testing HH. Wealth stratification among grooms' families within a given marriage market is a *necessary* condition for the possibility that footbinding leads to hypergamous marriage. This is to say that, if wealth across grooms' families is flat and every family has the same net worth, it is impossible that some brides distinguish themselves from others by marrying up the socio-economic ladder. (It is probable that this necessary condition is not adequately met in the context of the LMH data. Bossen and Gates testify to the widespread poverty of their research participants. See 2017, 44, 69.) However, it does not follow that wealth stratification among grooms' families is a *sufficient* condition for testing HH. For the evolutionary sciences approach, female mate competition occupies a crucial role in understanding footbinding's origins and maintenance. An open marriage market with female mate competition between natural-footed and footbound young women is also a necessary condition.

This raises the following methodological question. Why do Brown, Gates and Bossen modify their dataset by removing data in order to test HH only against birth cohorts where the vast majority of girls were footbound? They write, "The relationship hypothesized between footbinding and hypergamy was less likely once footbinding had become a minority practice than when footbinding was the norm. Therefore, we included only those birth cohorts in which the majority of women were bound" (Brown et al. 2012, 1048). Far from being charitable to HH, this mangled reasoning gets this relationship wrong-way round. From an evolutionary sciences approach, more competitive conditions involving footbound and natural-footed women are likely to catalyze greater rates of hypergamy among the footbound, *ceteris paribus*.

Driving confidence in their rejection of HH still lower is information pertaining to the one region in the modified footbinding data that had a slightly more open marriage market than others, Sichuan. These data differ in three ways from the other data. They were collected with a different survey instrument using different questions; collected by Gates 10 to 15 years before the group's supplemental data; and the number of Sichuan participants is many times larger than the number of participants in the other regions put together (according to available information, 7.46 times larger). According to Table 7 (Brown, et al., 2012, 1058), the Central region had 36 participants total (footbound and natural footed), the North region 261, and Southwest 142, for a grand total of 439 subjects. This is together merely 13% of the Sichuan total of 3,275. In the Sichuan data, *footbound girls frequently composed less than 90% of birth cohorts.* Sichuan villages sampled by Gates thereby fostered a more open, even if only slightly more open, marriage market in which footbound girls competed for wealthy grooms against non-footbound girls.

So did footbound girls in Sichuan attain hypergamous marriages at a rate higher than naturalfooted girls? They did indeed (Brown 2016, 515), as predicted by HH. The correlation between hypergamy and footboundedness among these participants found by the LMH team is probably significantly higher than reported. This is because the LMH team also conducted *post hoc* data reduction of the Sichuan data prior to testing so as to remove from analyses all data from cohorts of women wherein a majority was not footbound (see Brown 2016, 1048).

7 Does the LMH team appropriately operationalize the Hypergamy Hypothesis for testing?

Recall that advocates of LMH asked elderly women to remember the wealth differential between their natal and marriage families when they got married. Bossen, Gates and Brown write,

Footbinding & Hypergamy

 "we looked at three factors that were important indicators of rural wealth in order to represent relative economic standing in our rural communities prior to collectivization: ownership of land, a house, and a draft animal (usually an ox)" (2012, 1051). Original survey questions included:

a) How much land did your natal family have?

b) Did your natal family have a draft animal?

c) Did your natal family have a house?

d) How much land did your marriage family have?

e) Did your marriage family have a draft animal?

f) Did your marriage family have a house? (Brown, et al. 2012, 1051-2)

Computing the difference between these sets of answers about the natal and about the marriage family can provide information relevant for assessing HH so long as the following conditional is true: *if the groom's family possesses more net wealth than the bride's family, then that marriage is hypergamous*. A 'yes' answer is sufficient for a hypergamous marriage.

This straightforward, intuitive way of assessing marriage status leads to what we can call the 'Simple Procedure' for operationalizing the Hypergamy Hypothesis. On the Simple Procedure a marriage qualifies as hypergamous, in the context of data collected by the LMH team, if a bride answers 'Yes' to at least one of questions A, B or C:

A) (Arable acreage) Does the groom's family (i) own *more arable acreage* than the bride's family, (ii) own *as many or more* draft animals as the bride's family, *and* (iii) own *as many or more* houses as the bride's family?

B) (Draft animals) Does the groom's family (i) own *more draft animals* than the bride's family,
(ii) own *as much or more* arable acreage as the bride's family, *and* (iii) own *as many or more* houses as the bride's family?

C) (House) Does the groom's family (i) own *more houses* than the bride's family, (ii) own *as much or more* arable acreage as the bride's family, *and* (iii) own *as many or more* draft animals as the bride's family?

A, B and C carry a lot of logical structure. By doing so they represent single questions crafted to demarcate the explanation space by cycling through the three wealth variables in such a way that each question asks, of one and only one of the three resources, whether the groom's family had more of that resource than the bride's. This is done while holding the other variables constant.

The Simple Procedure proposes that a marriage is hypergamous so long as *at least one of A, B, or C is answered affirmatively.* This is for a compelling methodological reason. Survey participants tested by the LMH team were drawn from hardscrabble rural families trying to feed and clothe their children in shifting times of famine, war, currency devaluation, and political upheaval (Bossen & Gates 2017, 44). If the groom's family owns a house but the bride's family does not, then, in this social ecology, even if neither family has arable acreage or draft animals, the marriage is hypergamous. If all three questions receive a 'yes' from a participant, then the bride's marriage was extremely hypergamous.

Notice that to the survey questions listed above participants gave categorical answers. We reported that the LMH team modified the data from these survey questions prior to testing both by categorizing data by "birth cohorts," by altering the number of years governing what is a "birth cohort," and by removing data from participants who were in birth cohorts in which the majority was not footbound. Prior to testing HH, the LMH authors modified their data another way: *they transformed these categorical data into ordinal data.* They set land equal to 2 units of wealth, a house equal to 1, and a draft animal also equal to 1. What reason would impartial researchers have for such a modification? Authors discuss other modifications of the data. They mention reassigning draft animals as 1.5 units of wealth. In their own words, in the end, they "*arbitrarily* chose the first index,

where a house and draft animal are assigned equal value" (2012, 1055, my italics) but where land is weighted as more valuable.

The LMH team surprise readers on yet another occasion in their operationalization of HH. Rather than testing individual answers to the land, house, and animal questions with their modified wealth valuations, authors create and test an *ad hoc* aggregate variable they call the "marriage mobility index." Use of this index combines three different points of data, which have already been changed from categorical to ordinal data, from three independent categorical questions (house, land, animal). Granularity of data about wealth difference is lost in this transformation. On this occasion authors do not explain why collapsing these three variables was needed.

With these steps in place, Brown and colleagues calculate the "marriage mobility index" scores for those footbound and natural-footed women whose data remain. Authors compare these two averaged scores and find that they are not significantly different. Since HH predicted that footbound brides would marry up at greater rates than non-footbound brides, the team concludes HH is falsified.

This means of operationalizing HH is convoluted, unmotivated, and violates statistical and methodological best practices. For brevity consider only the "marriage mobility index." Participants were never asked to rank order the monetary value of animals, houses, and land. This index is the creation of Gates, Bossen and Brown. The revised data misrepresent survey responses. Authors offer no evidential justification from microeconomic studies of early twentieth-century China for any ordinal recoding of the market value of house, land, and draft animal. On top of this, they "arbitrarily" selected their recoding scale. By operationalizing HH only after modifying data in several ways, the LMH group artificially reduces the chance that their analyses would reveal a positive statistical relationship between hypergamy and footbinding.

Does the LMH team appropriately test the Hypergamy Hypothesis?

The LMH team purports to show that HH is false with a loglikelihood goodness-of-fit test. This test allows comparison of two different statistical models which, in this case, are HH and the null hypothesis. The null hypothesis affirms that footbound girls had no statistically larger correlation with hypergamous marriage than did natural-footed girls. HH is interpreted to imply that footbound brides were on average more likely to marry up than natural-footed brides. This sets our expected values.

In the minds of matchmakers, the quality of the feet of a prospective bride is represented as a continuous variable. In fact, matchmakers frequently took a sample shoe for a prospective girl's footbound feet to prospective grooms for review so that grooms could assess an honest representation of foot size. Popular ranking systems used to assess footbound feet classified feet at one of many ordinal positions (Gao 2007). Since the smallness of feet is likely to correlate with mobility, and less mobility may predict lower probabilities of cuckoldry in the minds of grooms, footsize appears relevant for (if not essential to) testing of HH. For its part, the LMH team make footbinding a categorical variable in their statistical tests. Subjects were categorized as either footbound or never bound. The team did not ask women how others had appraised their foot size at marriage, let alone collect information about foot or shoe size at marriage.

The LMH team then splits these two groups into three further groups based on their "marriage mobility index" score. If a woman's score was a positive number, she belongs in the hypergamous or upward group; if the score is zero, she belongs in the 'no mobility' group; if the score is a negative number, she goes into the hypogamous or downward group. Once these cells are filled, actual observed values are known. At this point, the loglikelihood statistic was used to compare the expected and observed values. Authors explain that "This test asks whether the proportions of women who have upward, downward, or no mobility are dependent on footbinding status" (2012, 1057). In the sample described in Table 7 at Brown, et al., 2012 (p. 1058), a total of 1,139 footbound women and 2,575 non-footbound women were represented for a grand total of 3,714 women. A simple, robust, and methodologically appropriate way to report a loglikelihood result would be to calculate the ratio of expected and observed values for all 1,139 footbound women across the three marriage categories, then do the same for the 2,575 unbound women. At this point we can test the differences between the ratios of hypergamous footbound women to all footbound women, and the ratios of hypergamous non-footbound women to all non-footbound women. Just as with the chi-square test statistic, if the ratios are not significantly different, or if they are but nonfootbound women are shown to marry hypergamously more than footbound women, then (overlooking other methodological problems) HH would be falsified.

The LMH team chooses not to calculate the ratio of expected and observed values for all women in their sample. Instead they split the 3,714 participants into multiple categories based on a *post hoc* variable of their own creation, "region," first encountered above in §6. With three hypergamy categories (up, down, neutral), two footbinding categories (bound/never bound), and four regions (Central, North, Sichuan, and Southwest), authors perform numerous statistical tests to compare each of a profusion of pairs of ratios.

This method of statistically testing HH issues in problems. First, were they to have important theoretical reasons to think that HH would apply differently across regions, authors should still report statistical test results for the entire sample of 3,714 people. No reason to expect significant regional variation in rates of footbinding is provided. No overall test is performed. Second, the LMH team reports no controls for familywise error. The threshold for statistical significance is never adjusted to factor for the increased rate of Type 1 error due to the numbers of statistical tests they run. Independent of these issues, a damning problem looms. Let's look closer at one region to get a sense for the cogency of their statistical test results disconfirming HH. In the "Central" region authors report that 17% of footbound women married up compared to 33% of non-footbound women; 50% of the footbound women married neither up nor down as compared to 43% of the non-footbound women; and 33% of footbound women married down as compared to 23% of non-footbound women. Authors infer these data from the Central region disconfirm HH. This is because this distribution of marriage type yields a non-significant p-value, calculated solely for the Central region, of "p = 0.68" (2012, 1058). The "Central" region represents one of only four total regions tested. Brown and the LMH team give the clear impression to readers that statistics from the Central region describe approximately a fourth of participants in the dataset. Amplifying this expectation, Brown, Gates and Bossen often use locutions such as "for most regions" (2012, 1035) and "across regions" (2012, 1057). The message is that any way readers look at their data, HH is disconfirmed.

This conceals a fundamental complication in the team's high-profile 2012 paper (Brown et al. 2012). Masked as it is, this problem is easy to miss for even diligent readers if they neglect appendices. The problem can be illustrated by the fact that *among the 3,714 total participants in the study, a grand total of six (6) were footbound and in the Central region.*

This knowledge permits a reappraisal of their earlier descriptive statistics. Readers are likely to be impressed when learning that twice as many natural-footed women from this region married up than did footbound women. So long as their readers missed the many modifications they did to their data prior to testing, this would appear to be a clear refutation of HH. What this claim actually means is that *two (2) natural-footed women in the "Central" region married up*. When they wrote that 16.66% of footbound women in the Central region married hypergamously, what they actually mean is that *one single footbound woman married hypergamously*. This amplifies concerns about using *region* rather than

village or simply *participant* as the unit of measurement to test HH, and about not testing the entire dataset as a whole. To repeat that "in most regions" there is no evidence for HH is disingenuous. The credibility gap widens when one realizes authors repeatedly treat the area with six total footbound women, whose data strongly disconfirmed HH, as evidentially equivalent to the region represented by 994 footbound women (Sichuan), which does confirm HH.

Measuring statistical significance forms less than half of the duty of social scientists when testing hypotheses. Brown and co-authors (Brown, et al., 2012) and Bossen and Gates (2017, 143, 144, 183, in neither in their logistic regression table (6.1), nor their odds ratio table (6.2), nor in their Appendix B about "Logistic Regression Results") report statistical power or effect sizes. It would be shocking if the tests of HH with Central region data had minimally acceptable effect sizes and statistical power, especially since sample size is a variable in the calculation of statistical power.

While concluding the above discussion about why an open marriage market is a necessary condition for a proper test of HH, we made two observations. (i) Data from Sichuan, collected earlier, represented a marriage market slightly more open than that described in other data. (ii) Data from Sichuan supported a positive statistical relationship between footbinding and hypergamous marriage. By coincidence, it is only when confronted with this apparent counterevidence to LMH from their Sichuan data that Brown and the LMH team voice interest in best practices about statistical significance testing. They write, "8 percent more women bound at marriage had marital households with houses than not-bound women." They then add, "That 8 percent difference is statistically significant, but is it important or meaningful? There is no definitive answer to this question, but the difference is slight..." (Brown, et al. 2012, 1057). *Reporting an effect size and statistical power would be valuable precisely in order to answer Brown's own question about these findings.* The difference to which Brown refers leads to a significance value of p = .03 (2012, 1054). When p values favored their preferred conclusion, that HH is false, skeptical questions like these were not raised.

9 What ought we conclude?

The purpose of this paper has been to subject the most important case against the Hypergamy Hypothesis for footbinding to a fair-minded critique that exemplifies common standards of the social sciences while being readable to researchers in the humanities. We have identified and explained in detail multiple methodological problems, statistical errors, repeated and unwarranted modifications of data, untrustworthy data analysis, and duplicitous reporting of test results from the LMH team's assessment of the Hypergamy Hypothesis. The evidence presented across several publications by the LMH team, consisting of Melissa J. Brown, Hill Gates, Laurel Bossen, and Damien Satterthwaite-Phillips, fails to disconfirm the Hypergamy Hypothesis.

This study has several implications. First, this points to the value of open science and opensource data. It is doubtful that Brown, Bossen, Gates and co-authors would have repeatedly and needlessly modified their data prior to testing had the data been publicly available in a repository. Second, this critical review sounds a cautionary note about humanities researchers' consumption of interdisciplinary research that employs statistical methods of analysis. The LMH team's flaunting of American Psychological Association standards for methods and statistical testing (as found in Brown et al. 2012) did not appear in a minor humanities journal with a limited circulation. It was published in the *Journal of Asian Studies*, the flagship journal of the Association for Asian Studies subscribed to by thousands of members.

A lingering, oblique factor contributing to the present ignorance about the likely origins, maintenance, and cessation of footbinding is the fact that the labor market explanation for footbinding faces no current competing account from the evolutionary social sciences. Dickemann's ground-breaking work was never centrally about footbinding; Dickemann's work was not informed by any emic knowledge of Chinese footbinding culture or customs; and this work appeared many decades ago. Design of a clear and thorough theoretical articulation of an evolutionary approach to

footbinding from the evolutionary social sciences could transform subsequent research on the practice. Such an account is likely to incorporate discussion of hypergamy, paternal certainty, mate guarding and other relevant principles, and, ideally, will exhibit knowledge about the actual history and practice of footbinding in China. Subsequent testing of such a theoretical model, whether with a dataset made publicly available or with the aid of an agent-based model, could dramatically increase our understanding of footbinding. Amplifying the potential value of this theoretical work are two final facts that, in combination, merit remedy. Discussion of footbinding occurs in nearly every introductory text for cultural anthropology, yet the dominant theory purporting to create understanding about footbinding suffers from dreadful problems.

Bibliography

- Blake, C. F. (1994). Foot-binding in neo-Confucian China and the appropriation of female labor. *Signs*, 676–712.
- Bossen, L. (2002). Chinese women and rural development: sixty years of change in Lu Village, Yunnan. Lanham, Md: Rowman & Littlefield Publishers.
- Bossen, L., & Gates, H. (2017). Bound feet, young hands: tracking the demise of footbinding in village China. Stanford, California: Stanford University Press.
- Bossen, L., Wang Xurui, Brown, M. J., & Gates, H. (2011). Feet and Fabrication: Footbinding and Early Twentieth-Century Rural Women's Labor in Shaanxi. *Modern China*, *37*(4), 347–383.
- Brown, M. J. (2016). Footbinding, Industrialization, and Evolutionary Explanation: An Empirical Illustration of Niche Construction and Social Inheritance. *Human Nature*, *27*(4), 501–532.
- Brown, M. J., Bossen, L., Gates, H., & Satterthwaite-Phillips, D. (2012). Marriage Mobility and Footbinding in Pre-1949 Rural China: A Reconsideration of Gender, Economics, and Meaning in Social Causation. *The Journal of Asian Studies*, 71(04), 1035–1067.

- Buss, D. M., & Shackelford, T. K. (1997). From vigilance to violence: mate retention tactics in married couples. *Journal of Personality and Social Psychology*, 72(2), 346–361.
- Chen, Y. (2014). The contested body: the anti-footbinding movement in modern China. Journal of Modern Chinese History, 8(2), 278–279. https://doi.org/10.1080/17535654.2014.960157
- Dickemann, M. (1979). The ecology of mating systems in hypergynous dowry societies. *Social Science Information*, 18(2), 163–195. <u>https://doi.org/10.1177/053901847901800201</u>
- Dickemann, M. (1991). Women, Class, and Dowry. *American Anthropologist*, *93*(4), 944–946. https://doi.org/10.1525/aa.1991.93.4.02a00110
- Dickemann, M. (1997). Paternal Confidence and Dowry Competition: A Biocultural Analysis of Purdah. In L. L. Betzig (Ed.), *Human nature: A critical reader* (pp. 311–328). Oxford University.
- Drucker, A. (1981). The Influence of Western Women on the Anti-Footbinding Movement 1840-1911. *Historical Reflections*, 8(3), 179–199.
- Fang, H. S. Y., & Yu, F. Y. K. (1960). Foot Binding in Chinese Women. Canadian Journal of Surgery. Journal Canadian de Chirurgie, 3, 195–202.
- Gates, H. (2008). Bound feet: How sexy were they? *The History of the Family*, *13*(1), 58–70. https://doi.org/10.1016/j.hisfam.2007.11.001
- Gates, H. (2017). Footbinding and Women's Labor in Sichuan. Routledge.
- Gao, Hongxing. 高洪兴. 2007.《缠足史》. 上海: 上海文艺出版社.
- Hartung, J., Dickemann, M., Melotti, U., Pospisil, L., Scott, E., Smith, J. M., & Wilder, W. (1982). Polygyny and Inheritance of Wealth. *Current Anthropology*, *23*(1), 1–12.
- Howard, J. A., & Gibson, M. A. (2017). Frequency-dependent female genital cutting behaviour confers evolutionary fitness benefits. *Nature Ecology & Evolution*, 1(3), 49.
- Ko, D. (2005). Cinderella's sisters: A revisionist history of footbinding. Univ of California Press.

Levy, H. S. (1966/1991). The lotus lovers: The complete history of the curious erotic custom of footbinding in *China*. Prometheus Books.

Mackie, G. (1996). Ending footbinding and infibulation: A convention account. *American Sociological Review*, 999–1017.

McMahon, D. (2014). Rethinking the Decline of China's Qing Dynasty: Imperial Activism and Borderland

Management at the Turn of the Nineteenth Century. Taylor and Francis.