Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis Over 20 Years in the Framingham Heart Study

[Posted as supplied by author]

SUPPLEMENTARY ONLINE MATERIAL FOR:

Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis Over 20 Years in the Framingham Heart Study

James H. Fowler[1], Nicholas A. Christakis[2]

[1]*Department of Political Science, University of California, San Diego, CA 92093, USA*

[2]*Department of Health Care Policy, Harvard Medical School, and Department of Sociology, Harvard University, Cambridge, MA 02138, USA*

Assembling the FHS Social Network Dataset

Here, we describe the source data we work with and the new network linkage data we have appended to it.

The Framingham Heart Study is a population-based, longitudinal, observational cohort study that was initiated in 1948 to prospectively investigate risk factors for cardiovascular disease. Since then, it has come to be composed of four separate but related cohort populations: (1) the "Original Cohort" enrolled in 1948 (N=5,209); (2) the "Offspring Cohort" (the children of the Original Cohort and spouses of the children) enrolled in 1971 (N=5,124); (3) the "Omni Cohort" enrolled in 1994 (N=508); and (4) the "Generation 3 Cohort" (the grandchildren of the Original Cohort) enrolled beginning in 2002 (N=4,095). The Original Cohort actually captured the majority of the adult residents of Framingham in 1948, and there was little refusal to participate. The Offspring Cohort included the majority of the living offspring of the Original Cohort in 1971, and their spouses. The supplementary, multi-ethnic Omni Cohort was initiated to reflect the increased diversity in Framingham since the inception of the Original Cohort; 508 participants, of whom 33% were Black, 49% Hispanic, and 18% Asian, attended the first Omni exam between 1994 and 1998 (only a small number of subjects from the Omni cohort appear in our network, as alters). For the Generation 3 Cohort, Offspring Cohort participants were asked to identify all their children and the children's spouses, and 4,095 subjects were enrolled beginning in 2002. Published reports provide details about sample composition and study design for all these cohorts.[1-3]

Continuous surveillance and serial examinations of these cohorts provide longitudinal data. All of the participants are personally examined by FHS physicians (or, for the small minority for whom this is not possible, evaluated by telephone) and watched continuously for outcomes. The Offspring study has collected information on health events and risk factors roughly every four years for over 30 years. The Original Cohort has data available for roughly every two years for 60 years. Importantly, even subjects who migrate out of the town of Framingham (to points throughout the U.S.) remain in the study and, remarkably, come back every few years to be examined and to complete survey forms; that is, there is no necessary loss to follow-up due to out-migration in this dataset, and very little loss to follow-up for any reason (e.g., only 10 cases out of 5,124 in the Offspring Cohort have been lost).

For the purposes of the analyses reported here, exam waves for the Original cohort were aligned with those of the Offspring cohort, so that all subjects were treated as having been examined at just seven waves (in the same time windows as the Offspring, as noted in Table S1).

The Offspring Cohort is the key cohort of interest here, and it is our source of "egos" (the focal individuals in our network). However, individuals to whom these egos are linked – in any of the four cohorts – are also included in the network. That is, whereas egos will come only from the Offspring Cohort, alters are drawn from the entire set of FHS cohorts (including also the Offspring Cohort itself). Hence, the total number of individuals in the FHS social network is 12,067, since alters identified in the Original, Generation 3, and Omni Cohorts are also included, so long as they were alive in 1971 or later.

The physical, laboratory, and survey examinations of the FHS participants provide a wide array of data. At each evaluation, subjects complete a battery of questionnaires (*e.g.*, the CES-D measure

of happiness and depression, as described below), a physician-administered medical history (including review of symptoms and hospitalizations), a physical examination administered by physicians on-site at the FHS facility, and a large variety of lab tests.

Table S1: Survey Waves and Sample Sizes of the Framingham Offspring Cohort (Network Egos)

Survey Wave/ Physical Exam	Time period	N alive	Number Alive and 18+	N examined	% of adults participating
Exam 1	1971-75	5124	4914	5,124	100.0
Exam 2	1979-82	5053	5037	3,863	76.7
Exam 3	1984-87	4974	4973	3,873	77.9
Exam 4	1987-90	4903	4903	4,019	82.0
Exam 5	1991-95	4793	4793	3,799	79.3
Exam 6	1996-98	4630	4630	3,532	76.3
Exam 7	1998-01	4486	4486	3,539	78.9

Table S1 provides information about the participation rates for each exam/survey wave. Given the size of the sample and the need to physically examine each participant at each survey wave, participants are examined on a rolling basis during windows of time, as indicated. Participant compliance with examinations is excellent, with each wave having a participation rate of about 80%.

Data collection and subject follow-up procedures at the FHS are superb. For example, the quality assurance protocol for physician examiners includes initial certification and annual retraining. Hospital and nursing home records and outside physician office records are routinely sought for all cardiovascular, fracture, and cancer events, and for all deaths.

To ascertain the network ties, we computerized information from archived, handwritten documents that had not previously been used for research purposes, namely, the administrative tracking sheets used and archived by the FHS since 1971 by personnel responsible for calling participants in order to arrange their periodic examinations. These tracking sheets were used as a way optimizing participant follow-up, by asking participants to identify people close to them. But they also implicitly contain valuable social network information. These sheets record the answers when all 5,124 of the egos were asked to comprehensively identify friends, neighbors (based on address), co-workers (based on place of employment), and relatives who might be in a position to know where the egos would be in two to four years. The key fact here that makes these administrative records so valuable for social network research is that, given the compact nature of the Framingham population in the period from 1971 to 2007, many of the nominated contacts were themselves also participants of one or another FHS cohort.

We have used these tracking sheets to develop network links for FHS Offspring participants to other participants in any of the four FHS cohorts. Thus, for example, it is possible to know which participants have a relationship (*e.g.*, spouse, sibling, friend, co-worker, neighbor) with other participants. Of note, each link between two people might be identified by *either party* identifying the other; this observation is most relevant to the "friend" link, as we can make this link either when A nominates B as a friend, or when B nominates A (and, as discussed below,

this directionality might also be substantively interesting). People in any of the FHS cohorts may marry or befriend or live next to or work with each other.

Finally, complete records of participants' and their contacts' addresses since 1971 are available. We have exploited this information as well, using address-mapping technologies. Because of the high quality of addresses in the FHS data, the compact nature of Framingham, and the wealth of information available about each subject's residential history, we have been able to correctly assign addresses to virtually all subjects. We can thus (1) determine who is whose neighbor, and (2) compute distances between individuals.[4]

Measuring Happiness: Factor Analysis of the Center for Epidemiological Studies Depression Scale (CES-D)

Table S2 shows results from a maximum-likelihood factor analysis fitting four factors to the CES-D. Factor loadings indicate how important each variable is for contributing to a latent factor that variables have in common. To focus on the most important variables for each factor, only loadings with a magnitude greater than 0.3 are shown. The results indicate that the first four questions in the table that we use to measure happiness are the best fitting variables for factor 3. Values on the loadings for factor 3 are negative because these four questions are worded positively (higher values indicate less depression), whereas all the other questions are worded negatively (higher values indicate more depression). These results confirm prior published work about the use of this scale to measure happiness, as described in the text.

Table S2: Factor Analysis of the Center for Epidemiological Studies Depression Scale (CES-D)

	Factor Loadings			
Question	*Factor 1*	*Factor 2*	***Factor 3***	*Factor 4*
I enjoyed life			**-0.727**	
I was happy			**-0.706**	
I felt hopeful about the future			**-0.573**	
I felt that I was just as good as other people			**-0.465**	
I felt that people disliked me				0.751
People were unfriendly				0.532
I thought my life had been a failure				0.302
I felt sad	0.714			
I had crying spells	0.625			
I felt depressed	0.596	0.465		
I felt that I could not shake off the blues	0.562	0.448		
I felt lonely	0.441	0.317		
I felt fearful	0.373	0.317		
I was bothered by things that usually don't bother me	0.344	0.347		
My sleep was restless		0.318		
I talked less than ususal		0.321		
I had trouble keeping my mind on what i was doing		0.493		
I felt that everything i did was an effort		0.648		
I did not feel like eating: my appetite was poor		0.334		
I could not "get going"		0.614		

Statistical Information and Sensitivity Analyses

This supplement contains tables of regression coefficients using the methods described in the main text. For the analyses in Tables S3, S4, and S5, we considered the prospective effect of alters, social network variables, and other control variables on future happiness. For the analyses in Tables S7, S8, S9, S10, and S11 we conducted regressions of ego happiness as a function of ego's age, gender, education, and happiness in the previous exam, and of the happiness of an alter in the current and previous exam. Inclusion of ego happiness at the previous exam eliminates serial correlation in the errors and also substantially controls for ego's genetic endowment and any intrinsic, stable predilection to be happy. Alter's happiness at the previous exam helps control for homophily.[5]

The key coefficient in these models that measures the effect of induction is on the variable for alter contemporaneous happiness.[6] We used generalized estimating equation (GEE) procedures to account for multiple observations of the same ego across waves and across ego-alter pairings.[7] We assumed an independent working correlation structure for the clusters.[8]

These analyses underlie the results presented in Figures 4 and 5 in the paper.

The GEE regression models in the tables provide parameter estimates in the form of beta coefficients, whereas the results reported in the text and in Figure 4 of the paper are in the form of risk ratios, which are related to the exponentiated coefficients. Mean effect sizes and 95% confidence intervals were calculated by simulating first difference in alter contemporaneous happiness (changing from 0 to 1) using 1,000 randomly drawn sets of estimates from the coefficient covariance matrix and assuming all other variables are held at their means.[9]

The regression coefficients have mostly the expected effects, such that, for example, ego's previous happiness is the strongest predictor for current happiness. The models in the tables include exam fixed effects, which, combined with age at baseline, account for the aging of the population. The sample size is shown for each model, reflecting the total number of all such ties, with multiple observations for each tie if it was observed in more than one exam, and allowing for the possibility that a given person can have multiple ties.

We evaluated the possibility of omitted variables or contemporaneous events explaining the associations by examining how the type or direction of the social relationship between ego and alter affects the association between ego and alter. If unobserved factors drive the association between ego and alter friendship, then directionality of friendship should not be relevant.

We explored the sensitivity of our results to model specification by conducting numerous other analyses (not shown here) each of which had various strengths and limitations, but none of which yielded substantially different results than those presented here. For example, we used the raw scaled happiness score as a continuous variable in an ordinary least squares model and found the same significant relationships. We also experimented with different error specifications. Although we identified only a single friend for most of the egos, we studied how multiple observations on some egos affected the standard errors of our models. Huber-White sandwich estimates with clustering on the egos yielded very similar results. We also tested for the presence of serial correlation in the GEE models using a Lagrange multiplier test and found none remaining after including the lagged dependent variable.[10]

We also considered the possibility that clustering in education or income may drive our results. Similarity in socioeconomic status probably cannot explain the clustering of happy people since next-door neighbors have a much stronger influence than neighbors who live a few doors down in the same neighborhood (and who consequently have similar housing wealth and environmental exposures). Nor do we find evidence of geographic clustering at a larger scale. The maps in Figure S1 further show that the geographic distribution of happiness is not systematically related to local levels of either income or education.

Network Analysis

To be sure the clustering of happy and unhappy people in Figure 1 of the main text was not simply due to chance, we compared the observed mean cluster size to the mean cluster size in 1,000 randomly generated networks in which we preserved network topology and the overall prevalence of happiness but randomly shuffled the assignment of the happiness value to each node. This procedure indicates that clusters of connected happy individuals are significantly larger in the observed network in both 1996 (+0.10 nodes, 95% C.I. 0.03-0.17) and 2000 (+0.19 nodes, 95% C.I. 0.10-0.26).

The Kamada-Kawai algorithm used to prepare the images in Figure 1 of the main text generates a matrix of shortest network path distances from each node to all other nodes in the network and repositions nodes so as to reduce the sum of the difference between the plotted distances and the network distances.[11]

Eigenvector centrality assumes that the centrality of a given individual is an increasing function of the centralities of all the individuals to whom he or she is connected.[12] While this is an intuitive way to think about which subjects might be better connected, it yields a practical problem: how do we simultaneously estimate the centrality of all subjects in the network? Let a_{ij} equal 1 if subjects *i* and *j* have a social connection and 0 if they do not. Furthermore, let $\boldsymbol{x}$ be a vector of centrality scores so that each subject's centrality x_j is proportional to the sum of the centralities of the subjects to whom they are connected: $\lambda x_i = a_{1i}x_1 + a_{2i}x_2 + \cdots + a_{ni}x_n$. This yields *n* equations, which can be represented as $\lambda x = A^T x$. The vector of centralities $\boldsymbol{x}$ can now be computed since it is an eigenvector of the eigenvalue λ. Although there are *n* nonzero solutions to this set of equations, in symmetric matrices, the eigenvector corresponding to the principal eigenvalue is used because it maximizes the accuracy with which the associated eigenvector can reproduce the original social network.[13] To be sure of reaching a solution, we symmetrized all asymmetric relationships in the observed network (i.e., we assumed all friendship ties were mutual).

Table S3. Aggregate Influence of Alters on Future Ego Happiness

	Model 1			Model 2			Model 3		
	Coef	**S.E.**	***p***	**Coef**	**S.E.**	***p***	**Coef**	**S.E.**	***p***
Number of Happy Alters	0.09	0.03	0.00						
Number of Unhappy Alters	-0.07	0.02	0.00						
Number of Alters				0.00	0.02	0.97			
Fraction of Alters Who Are Happy							0.28	0.08	0.00
Happiness in Previous Exam	1.32	0.07	0.00	1.33	0.07	0.00	1.32	0.07	0.00
Exam Number	-0.69	0.08	0.00	-0.56	0.07	0.00	-0.65	0.07	0.00
Constant	4.41	0.44	0.00	3.55	0.43	0.00	4.05	0.45	0.00
Deviance	1124			1128			1125		
Null Deviance	1230			1230			1230		
N	5261			4872			5261		

Table S4. Aggregate Influence of Alters on Future Ego Happiness, with Controls

	Coef.	**S.E.**	***p***
Number of Happy Alters	0.12	0.04	0.00
Number of Unhappy Alters	-0.06	0.03	0.08
Fraction of Alters who are Happy	-0.07	0.14	0.62
Happiness in Previous Exam	1.24	0.07	0.00
Age	0.00	0.00	0.54
Education	0.07	0.01	0.00
Female	-0.13	0.06	0.04
Exam Number	-0.71	0.08	0.00
Constant	3.49	0.57	0.00
Deviance	1052		
Null Deviance	1151		
N	4909		

Results for logistic regression of ego happiness at next exam (1=happy, 0=isn't happy) on covariates are shown in first column of Tables S3 and S4. Models were estimated using a general estimating equation (GEE) with clustering on the ego and an independent working covariance structure.[7,8] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[14] The results show that number of happy alters is the best predictor of future ego happiness.

Table S5. Influence of Ego Centrality on Future Ego Happiness

	Simple Model				*Model with Controls*			
	Coef.	**S.E.**	**Wald**	**p(>W)**	**Coef.**	**S.E.**	**Wald**	**p(>W)**
Centrality	5.49	2.49	4.85	0.03	5.22	2.61	4.00	0.05
Happiness in Previous Exam	1.29	0.06	441.46	0.00	1.25	0.06	374.14	0.00
Non-family Alters in Previous Exam					0.13	0.03	16.50	0.00
Family Alters in Previous Exam					-0.02	0.01	3.49	0.06
Age					-0.01	0.00	12.10	0.00
Education					0.08	0.01	36.78	0.00
Female					-0.16	0.06	7.68	0.01
Exam Number	-0.48	0.06	69.16	0.00	-0.49	0.06	66.39	0.00
Constant	2.95	0.36	65.61	0.00	2.61	0.45	33.23	0.00
Deviance	1454				1320			
Null Deviance	1567				1461			
N	6573				6113			

Results for logistic regression of ego happiness at next exam (1=happy, 0=isn't happy) on covariates are shown in first column. Models were estimated using a general estimating equation (GEE) with clustering on the ego and an independent working covariance structure.[7,8] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[14] The results show that eigenvector centrality is a significant predictor of future ego happiness.

Table S6a: Association of Alter Happiness and Ego Happiness

	Alter Type					
	Nearby Friend	Distant Friend	Nearby Mutual Friend	Nearby Alter-Perceived Friend	Coresident Spouse	Non Coresident Spouse
Alter Currently Happy	0.70	-0.10	2.07	0.32	0.21	0.05
	(0.34)	(0.21)	(0.79)	(0.41)	(0.11)	(0.32)
Alter Previously Happy	-0.21	0.60	-1.87	0.46	0.11	0.27
	(0.37)	(0.22)	(0.90)	(0.40)	(0.11)	(0.30)
Ego Previously Happy	1.89	1.27	3.19	1.46	1.36	1.34
	(0.40)	(0.24)	(0.99)	(0.44)	(0.12)	(0.26)
Exam 7	-0.73	-0.71	-1.49	-0.86	-0.81	-0.42
	(0.40)	(0.23)	(1.04)	(0.42)	(0.12)	(0.29)
Ego's Age	-0.04	0.01	-0.08	-0.01	0.01	0.04
	(0.02)	(0.01)	(0.03)	(0.02)	(0.01)	(0.02)
Ego Female	-0.53	-0.30	-0.74	-0.60	-0.29	-0.20
	(0.33)	(0.22)	(0.53)	(0.43)	(0.10)	(0.25)
Ego's Years of Education	0.01	0.15	-0.06	-0.02	0.04	0.12
	(0.08)	(0.05)	(0.13)	(0.10)	(0.02)	(0.06)
Constant	2.62	-2.14	6.49	1.49	-0.58	-3.90
	(1.75)	(1.30)	(3.00)	(2.38)	(0.58)	(1.45)
Deviance	46	104	14	29	417	64
Null Deviance	56	117	22	33	462	73
N	258	521	114	153	2018	307

Coefficients and standard errors in parenthesis for logistic regression of ego happiness (1=happy, 0=isn't happy) on covariates are shown. Observations for each model are restricted by type of relationship (*e.g.*, the leftmost model includes only observations in which the ego named the alter as a "friend" in the previous and current period, and the friend is "nearby" – i.e. lives no more than 1 mile away). Models were estimated using a general estimating equation with clustering on the ego and an independent working covariance structure.[7,8] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[14]

Table S6b: Association of Alter Happiness and Ego Happiness

	Alter Type					
	Nearby Sibling	Distant Sibling	Immediate Neighbor	Neighbor within 25M	Neighbor within 100M	Co-worker
Alter Currently Happy	0.32	0.05	0.83	0.10	-0.11	-0.29
	(0.15)	(0.08)	(0.31)	(0.16)	(0.08)	(0.16)
Alter Previously Happy	0.02	0.14	0.30	0.01	-0.01	-0.13
	(0.15)	(0.09)	(0.37)	(0.15)	(0.08)	(0.22)
Ego Previously Happy	1.67	1.35	1.35	1.15	1.30	1.46
	(0.20)	(0.14)	(0.55)	(0.28)	(0.17)	(0.52)
Exam 7	-0.84	-0.71	-0.66	-0.24	-0.63	-0.89
	(0.20)	(0.13)	(0.52)	(0.28)	(0.17)	(0.46)
Ego's Age	0.01	0.02	-0.02	0.02	0.01	-0.01
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
Ego Female	0.18	-0.11	-0.10	-0.23	-0.09	-0.12
	(0.16)	(0.11)	(0.46)	(0.24)	(0.14)	(0.38)
Ego's Years of Education	0.06	0.07	0.10	0.07	0.06	0.05
	(0.04)	(0.03)	(0.12)	(0.05)	(0.03)	(0.09)
Constant	-1.23	-1.69	0.02	-2.09	-1.18	0.84
	(0.84)	(0.62)	(2.25)	(1.19)	(0.73)	(1.66)
Deviance	232	703	35	205	755	122
Null Deviance	269	778	42	221	821	135
N	1117	3297	186	965	3496	600

Coefficients and standard errors in parenthesis for logistic regression of ego happiness (1=happy, 0=isn't happy) on covariates are shown. Observations for each model are restricted by type of relationship (*e.g.*, the leftmost model includes only observations in which the ego named the alter as a "sibling" in the previous and current period, and the sibling is "nearby" – i.e. lives no more than 1 mile away). Models were estimated using a general estimating equation with clustering on the ego and an independent working covariance structure.[7,8] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[14]

Measures of Occupational Prestige

The Framingham dataset does not itself contain specific occupational information. However, we were able to construct a measure of occupational prestige by using occupation data obtained from tracking records used by the study administrators but not previously used for research, and also data obtained from public records in Framingham and adjoining towns (as part of New England town Censuses).

This data was then coded using the International Standard Classification of Occupations (ISCO-88). Occupations coded in this way can be easily recoded into various other scales using freely available software.[15]

Individuals were assumed to keep their occupation from the date recorded at a particular wave until the next change. Where waves were missing, the previous code was entered if the same occupation was measured again at a later date.

Unfortunately, it was not possible to code occupations for all subjects at all waves. The table below gives the rates of available information. A total of 80% of the people have occupational prestige scores available for at least one wave.

Table S7: Availability of Occupational Prestige Data

Data Wave	Year	% Coded	% Coded (Incl. Married Women)	Mean Treiman Score (NIC Married Women)
1	1973	42	56	47
2	1979	58	58	47
3	1987	56	63	48
4	1991	53	59	48
5	1993	46	50	49
6	1998	38	42	49
7	2000	34	37	49

Once occupations have been assigned ISCO-88 codes, the occupations can then be mapped to occupational prestige scores using a variety of extant methods. Here, occupational prestige is coded as a Treiman score, which places occupations in an ordered scale based on public perceptions of their prestige. The scale runs hierarchically from 13 to 78.[16]

A difficulty with this is the assignment of prestige to married women. One possibility is to assign married women who are not listed with their own occupation the prestige scores of their husbands (a not unreasonable assumption give the time and place of the Framingham Offspring Cohort). Another option is to assign married women only the prestige of their own occupation and to code them as missing if "unemployed." When we add these variables to the nearby friend models, as shown in Table S8, neither approach yields a significant relationship between occupational prestige and happiness. The reason is that occupational prestige correlates strongly with education (ρ=0.51), which appears to be a superior proxy for socioeconomic status and its influence on happiness.

Table S8. Happiness Spreads Between Nearby Friends, Even When Controlling For Occupational Prestige.

	Unemployed Married Women Take Own Value			Unemployed Married Women Take Husband's Value		
	Coef.	*S.E.*	*p*	*Coef.*	*S.E.*	*p*
Alter Currently Happy	0.908	0.255	0.000	0.908	0.255	0.000
Alter Previously Happy	-0.120	0.234	0.609	-0.120	0.233	0.607
Ego Previously Happy	1.724	0.424	0.000	1.724	0.422	0.000
Exam 7	0.010	0.292	0.972	0.011	0.292	0.969
Ego's Age	0.024	0.006	0.000	0.024	0.006	0.000
Ego Female	-0.001	0.137	0.996	-0.001	0.136	0.995
Ego's Years of Education	0.078	0.034	0.021	0.078	0.034	0.020
Occupational Prestige	-0.004	0.006	0.496	-0.004	0.006	0.491
Constant	-3.420	0.844	0.000	-3.423	0.843	0.000
Deviance	1488.6			1488.6		
Null Deviance	1905.5			1905.5		
N	1511			1511		

Table S9: Association of Alter Happiness and Ego Happiness, by Physical Distance Between Ego and Alter

	Alter Type				
	Friend <0.5 miles	Friend <1 miles	Friend <2 miles	Friend <5 miles	Friend <10 miles
Alter Currently Happy	1.01	0.70	0.57	0.32	0.23
	(0.42)	(0.34)	(0.26)	(0.21)	(0.19)
Alter Previously Happy	-0.29	-0.21	-0.07	0.24	0.35
	(0.47)	(0.37)	(0.28)	(0.22)	(0.20)
Ego Previously Happy	1.72	1.89	1.57	1.44	1.28
	(0.49)	(0.40)	(0.29)	(0.23)	(0.22)
Exam 7	-0.46	-0.73	-0.61	-0.69	-0.67
	(0.48)	(0.40)	(0.28)	(0.23)	(0.21)
Ego's Age	-0.03	-0.04	-0.02	-0.01	0.00
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
Ego Female	-0.43	-0.53	-0.48	-0.46	-0.37
	(0.39)	(0.33)	(0.24)	(0.22)	(0.20)
Ego's Years of Education	-0.04	0.01	0.05	0.09	0.10
	(0.09)	(0.08)	(0.06)	(0.05)	(0.05)
Constant	2.59	2.62	0.85	-0.42	-0.71
	(2.22)	(1.75)	(1.30)	(1.19)	(1.13)
Deviance	32	46	81	114	134
Null Deviance	38	56	94	131	150
N	175	258	418	594	677

Coefficients and standard errors in parenthesis for logistic regression of ego happiness (1=happy, 0=isn't happy) on covariates are shown. Observations for each model are restricted by type of relationship (*e.g.*, the leftmost model includes only observations in which the ego named the alter as a "friend" in the previous and current period, and the friend lives no more than 0.5 miles away). Models were estimated using a general estimating equation with clustering on the ego and an independent working covariance structure.[7,8] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[14]

Table S10: Association of Alter Happiness and Ego Happiness, by Time Since Alter Happiness was Measured

	Alter Type				
	Friend <0.5 Years	Friend <1.0 Years	Friend <1.5 Years	Friend <2.0 Years	Friend <2.5 Years
Alter Currently Happy	1.34	1.01	0.53	0.29	0.23
	(0.58)	(0.40)	(0.31)	(0.28)	(0.28)
Alter Previously Happy	-0.26	-0.03	-0.07	0.13	0.17
	(0.56)	(0.45)	(0.34)	(0.31)	(0.30)
Ego Previously Happy	2.60	1.39	1.33	1.21	1.24
	(0.65)	(0.47)	(0.37)	(0.33)	(0.30)
Exam 7	-1.29	-0.44	-0.37	-0.52	-0.46
	(0.61)	(0.41)	(0.33)	(0.30)	(0.28)
Ego's Age	0.01	0.00	-0.01	-0.02	-0.01
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Ego Female	0.02	-0.61	-0.59	-0.49	-0.41
	(0.53)	(0.40)	(0.34)	(0.31)	(0.30)
Ego's Years of Education	0.06	0.09	0.15	0.12	0.11
	(0.12)	(0.09)	(0.08)	(0.07)	(0.07)
Constant	-1.63	-0.75	-0.58	-0.09	-0.11
	(2.88)	(2.00)	(1.86)	(1.77)	(1.70)
Deviance	17	34	48	58	62
Null Deviance	22	40	57	66	70
N	116	196	265	299	315

Coefficients and standard errors in parenthesis for logistic regression of ego happiness (1=happy, 0=isn't happy) on covariates are shown. Observations for each model are restricted by type of relationship (*e.g.*, the leftmost model includes only observations in which the ego named the alter as a "friend" in the previous and current period, and the friend's exam occurred no more than 0.5 years prior to the ego's exam). Observations restricted to friends that live within 5 miles. Models were estimated using a general estimating equation with clustering on the ego and an independent working covariance structure.[7,8] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[14]

Table S11. Influence of Opposite Gender Relations on Association Between Alter Happiness and Ego Happiness

	Coef.	S.E.	*p*
Alter is Opposite Gender * Alter Currently Happy	-0.40	0.17	0.02
Alter Currently Happy	0.62	0.16	0.00
Alter is Opposite Gender	0.25	0.13	0.06
Alter Previously Happy	0.06	0.08	0.44
Ego Previously Happy	1.50	0.11	0.00
Exam 7	-0.82	0.11	0.00
Ego's Age	0.00	0.00	0.31
Ego's Gender	-0.13	0.09	0.16
Ego's Education	0.05	0.02	0.02
Constant	-1.01	0.48	0.03
Deviance	748		
Null Deviance	845		
N	3647		

Results for logistic regression of ego happiness at next exam (1=happy, 0=isn't happy) on covariates are shown. Sample includes all next-door neighbors and nearby friends and siblings (nearby = less than a mile away). We restricted this analysis to nearby people given the importance of distance documented elsewhere in our results. The interaction term in the first row tests the hypothesis that alters of opposite gender have less influence than alters of the same gender. Models were estimated using a general estimating equation (GEE) with clustering on the ego and an independent working covariance structure.[7,8] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[14] The results show that opposite-sex pairs exert significantly less influence on each other than same-sex pairs, helping to explain why spouses exert less influence than friends and neighbors, as shown in Table S6.

Comparison of Spousal Effect on Happiness with Prior Studies

The literature on inter-spousal correlation of wellbeing involves many different measures, samples, time frames, and analyses and shows a robust correlation of the well being of one spouse with that of the other.[17] For example, in a study of 1,040 community-dwelling elderly, Bookwala and Shulz find that husbands had an average of 3.4 (SD 3.7) depressive symptoms on a 10-point scale, and wives an average of 4.6 (SD 4.4).[18] Cross-sectionally, a one-point increase in wife's depressive symptoms was associated with a 0.2 point increase in the husband's depressive symptoms. The effect of husbands on wives was identical. There are at least a dozen similar studies in the literature involving cross-sectional work, but these all suffer from a common problem: they very likely overstate the influence of one spouse on another because spouses probably choose one another in part based on their tendency to be depressed or the variables that underlie that tendency. Cross-sectional models cannot separate influence from homophily.

In contrast, the models presented here are longitudinal, and they control for the baseline happiness of both spouses and have repeated measures across time. A similar effort, by Siegel et al, used two waves of the HRS (N=5,035 respondents) separated by two years to model ego depressive symptoms (using a subset of the CESD) at time $t+1$.[19] The key predictors were the spouse's depressive symptoms at time t and the *change* in spousal depressive symptoms between times $t+1$ and t; also in the model were the ego's depressive symptoms at time t. These investigators made a number of idiosyncratic decisions about how to transform the variables measuring depression, but it appears that a 1-point increase in depressive symptoms (on an 8 point scale) in a spouse is associated with a 0.16 point increase in depressive symptoms in the ego.

The results presented here are in keeping with the foregoing in terms of rough magnitude. We find that a person who becomes happy increases the *probability* their spouse will be happy by 0.08 (95% CI: 0.002 to 0.16). For comparability with the Siegel et al paper, we also ran another model where we restricted observations to those in which the time between spouse's exams was two years or less (see Table S12). That model shows that the effect on probability increases to 0.12 (95% CI: 0.02 to 0.23). We also examined the effect in a continuous variable model (see Table S12) and found that a one point increase in a person's happiness index increases the spouse's happiness index by 0.14 points (95% CI: 0.06 to 0.23).

Table S12. Dichotomous and Continuous Models for Spouses, Restricting Observations to Two Year Separation Between Ego and Alter Exams

	Dichotomous Model (Logit)			Continuous Model (OLS)		
	Coef.	*S.E.*	*p*	*Coef.*	*S.E.*	*p*
Alter Currently Happy	**0.331**	**0.143**	**0.021**	**0.143**	**0.043**	**0.001**
Alter Previously Happy	0.069	0.150	0.644	0.008	0.029	0.787
Ego Previously Happy	1.083	0.163	0.000	0.200	0.033	0.000
Exam 7	-0.801	0.174	0.000	-0.695	0.147	0.000
Ego's Age	0.008	0.009	0.398	0.008	0.008	0.369
Ego Female	-0.340	0.142	0.016	-0.268	0.128	0.036
Ego's Years of Education	0.022	0.031	0.491	0.045	0.027	0.088
Constant	-0.241	0.805	0.765	6.631	0.881	0.000
Deviance	214			3948		
Null Deviance	232			4389		
N	1052			1052		

Logistic regression of ego happiness (1=happy, 0=isn't happy) and ordinary regression of ego happiness (happiness scale from 0=least happy to 12=most happy) are shown. Observations for each model are restricted to spouses that have had their current exam within 2 years of one another. Models were estimated using a general estimating equation with clustering on the ego and an independent working covariance structure.[7,8] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[14]

Are Spousal Effects Significantly Different from Friend Effects?

While the coefficient for the effect of friends is larger than the coefficient for the effect of spouses, we cannot statistically distinguish these effects. To conduct a specific test of the difference, we included friends and spouses in the same model and included an interaction term on alter's happiness to test the hypothesis that friends affect each other more than spouses (see Table S13). The coefficient on this term is not significant (p=0.26).

An important observation is that our models control for baseline traits, and this captures a variety of factors including a tendency to homophily. It is well-known that spouses are more alike in age, education, income, and in other traits that influence happiness than friends are (though friends too, obviously, evince some homophily). Therefore, controlling for homophily will attenuate the inter-spousal effect more than the inter-friend effect. The reason we do this is because we want to know what the influence effect is, *net of homophily*.

However, since intuitions about how much spouses affect each other are likely based on the combined effect of influence and homophily, controlling for homophily will naturally tend to produce measured estimates that may not comport with the magnitude our intuition suggests based on anecdotal evidence. Some support for this idea in our data comes from raw Pearson correlations in change in happiness. Among spouses, the correlation is 0.13 (95% CI: 0.09 to 0.17) and among friends it is 0.09 (95% CI: 0.02 to 0.16). When we do not control for homophily or other factors, spouses do appear to change together more often. But the difference in these estimates is not significant and the relationship reverses when we include adequate controls in the full models in Table S6a.

Finally, in our data, spouses are concordant about 60% of the time vs. friends who are concordant about 58% of the time. Compare that to random concordance, which would be 52% for the observed incidence of happiness. So spouses do exhibit higher concordance but not much higher. This suggests that any effect of error in measurement on our results would have about the same effect on spouses and friends, and thus it cannot explain the (non-significant) difference in point estimates.

Table S13. Model of Spouses and Friends Showing No Statistical Difference in Effect Size Between the Two Groups

	Coef.	*S.E.*	*p*
Alter Currently Happy * Alter is Spouse	**-0.426**	**0.381**	**0.263**
Alter Currently Happy	0.762	0.365	0.037
Alter Previously Happy	0.096	0.114	0.402
Ego Previously Happy	1.325	0.131	0.000
Exam 7	-0.783	0.134	0.000
Ego's Age	0.008	0.008	0.313
Ego Female	-0.341	0.115	0.003
Ego's Years of Education	0.032	0.025	0.209
Alter is Spouse (not Friend)	0.173	0.296	0.558
Constant	-0.701	0.689	0.309
Deviance	463		
Null Deviance	518		
N	2276		

Logistic regression of ego happiness (1=happy, 0=isn't happy) among spouses and friends, with an interaction term for the type of friendship. Models were estimated using a general estimating equation with clustering on the ego and an independent working covariance structure.[7,8] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[14]

Does Gender Help to Explain the Difference in Effect Size Between Friends and Spouses?

While happiness measures do not appear to be gendered, there are several suggestions in the literature that inter-personal effects *per se* are indeed gendered, and that men's moods/happiness/well-being affect women more than the other way around.[18,20-22] In our data, we do not find such an asymmetry. In our model of spouses, we added a term that interacts gender with the effect of alter on ego, and we do not find a significant difference between husbands and wives (see Table S14).

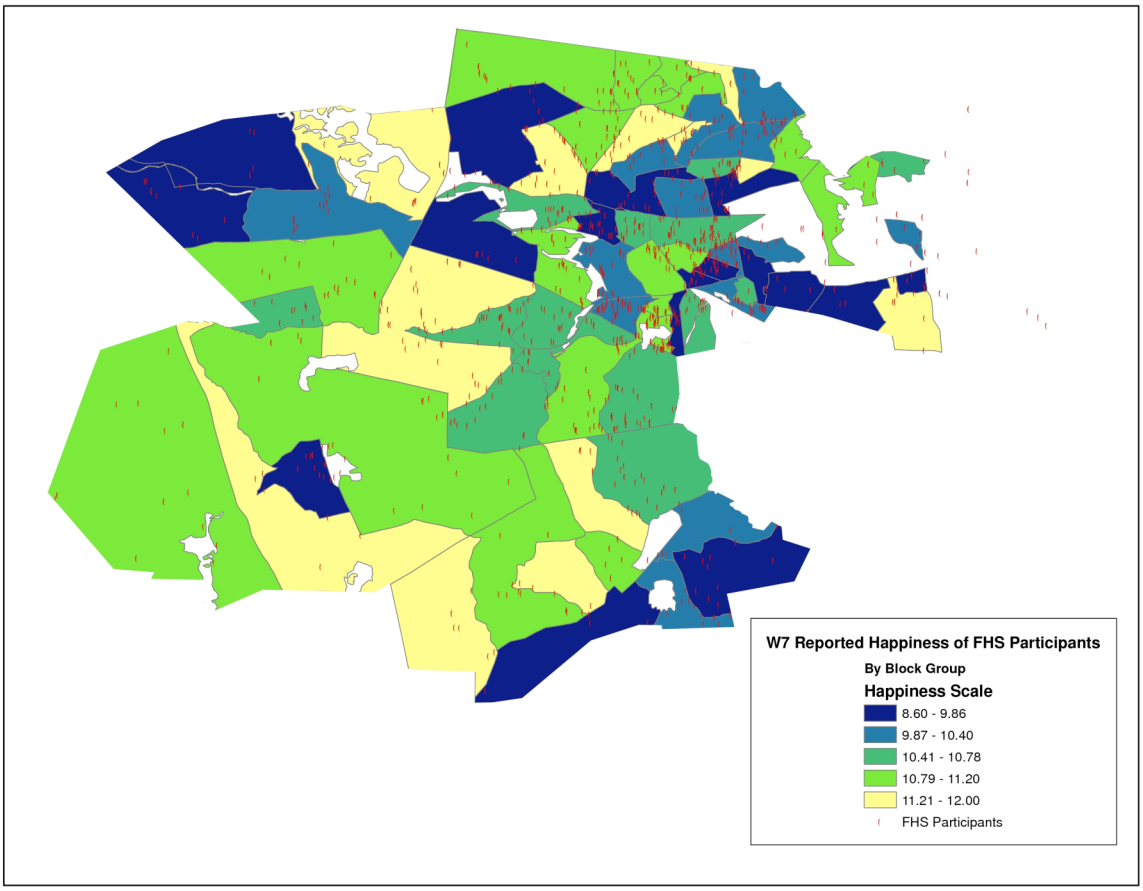
However, as already discussed in the paper, we did find that same-gender pairs of friends, siblings, and neighbors affect each other significantly more than opposite-gender pairs (p=0.02). Thus, the fact that most friends are of concordant gender and that all spouses are of discordant gender may help explain why we get a higher (though not significantly different) point estimate for friends than for spouses.

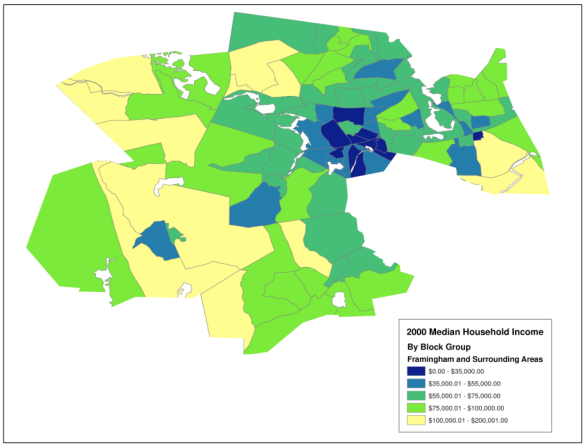
Table S14. Women are Not Significantly More Sensitive Than Men to Their Spouse's Happiness

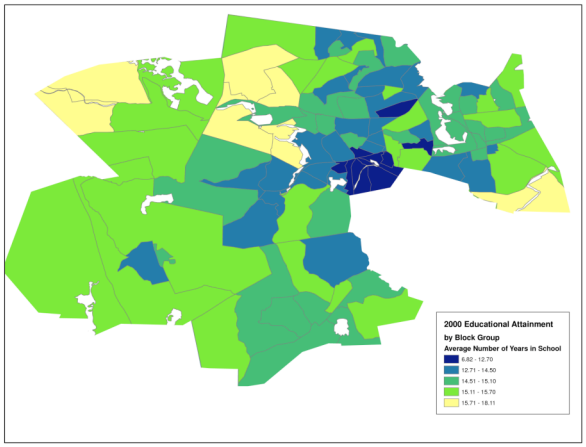
	Coef.	*S.E.*	*p*
Alter Currently Happy * Ego is Female	**0.037**	**0.228**	**0.871**
Alter Currently Happy	0.308	0.165	0.062
Alter Previously Happy	0.124	0.120	0.301
Ego Previously Happy	1.278	0.134	0.000
Exam 7	-0.775	0.136	0.000
Ego's Age	0.015	0.008	0.061
Ego Female	-0.347	0.187	0.063
Ego's Years of Education	0.031	0.026	0.223
Constant	-0.948	0.667	0.155
Deviance	417		
Null Deviance	462		
N	2018		

Logistic regression of ego happiness (1=happy, 0=isn't happy) among all spouses, with an interaction term for the gender of the ego. Models were estimated using a general estimating equation with clustering on the ego and an independent working covariance structure.[7,8] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[14]

Figure S1. Geographic Distribution of Happiness, Income, and Education in the Framingham Social Network







These figures show that the average happiness level (top) does not correspond to median income (lower left), or mean education level (lower right) for each block group in the Township of Framingham and the adjoining town of Natick. A Census Block Group must have at least 5 FHS subjects to be colored in happiness figure. Figures were drawn using ArcView. Note that these maps show only the town of Framingham and the adjoining town of Natick. While the greatest fraction of participants resided in these towns, participants were of course dispersed throughout the U.S. and our analyses made use of all cases, whatever their location.

References

1. Kannel WB, Feinleib M, McNamara PM et al. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol*. 1979;110:281-290.

2. Cupples LA, D'Agostino RB. Survival following initial cardiovascular events: 30 year follow-up. In: Kannel WB, Wolf PA, Garrison RJ, editors. The Framingham Study: An epidemiological investigation of cardiovascular disease. Bethesda, MD: NHLBI, NIH, 1988.

3. Quan SF, Howard BV, Iber C et al. The Sleep Heart Health Study: design, rationale, and methods. *Sleep*. 1997;20:1077-1085.

4. Fitzpatrick, Gary L., and Marilyn J. Modlin. *Direct-Line Distances: International Edition*. Metuchen, N.J.: The Scarecrow Press, 1986.

5. Carrington PJ, Scott J, and Wasserman S (2005) *Models and Methods in Social Network Analysis*. Cambridge University Press.

6. Christakis NA, Fowler JH (2007) The Spread of Obesity in a Large Social Network Over 32 Years. *New England Journal of Medicine* 357: 370-379.

7. Liang, K.Y. and Zeger, S.L. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; 73: 13-22.

8. Schildcrout JS. Regression analysis of longitudinal binary data with time-dependent environmental covariates: bias and efficiency. *Biostatistics* 2005 6(4):633-652.

9. King G, Tomz M, and Wittenberg J. Making the most of statistical analyses: improving interpretation and presentation. *American Journal of Political Science* 2000; 44: 341-355

10. Beck N. Time-Series–Cross-Section Data: What Have We Learned in the Past Few Years? *Annual Review of Political Science* 2001; 4: 271-293

11. Kamada T and Kawai S. An algorithm for drawing general undirected graphs. *Information Processing Letters* 1989; 31: 7-15.

12. Bonacich P. 1972. "Factoring and Weighing Approaches to Clique Identification." *Journal of Mathematical Sociology* 2:113-120.

13. Bonacich P. 1987. "Power and centrality: a family of measures." *American Journal of Sociology* 92(March): 1170-1182.

14. Wei P (2002) Goodness-of-fit Tests for GEE with Correlated Binary Data. *Scandinavian Journal of Statistics* 29 (1): 101110

15. Hendrick, J. (2002) ISKO: Stata module to recode 4 digit ISCO-88 occupational codes. Statistical Software Components S425802, Boston College Department of Economics.

16. Treiman, D.J. (1977) Occupational Prestige in Comparative Perspective. New York: Academic Press.

17 Meyler D., Stimpson JP, and Peek MK. Health Concordance Within Couples: A Systematic Review. Social Science and Medicine 2007; 64: 2297-2310.

18 Bookwala J and Schulz R. Spousal Similarity in Subjective Well Being: The Cardiovascular Health Study. Psychology and Aging 1996; 11: 582-590.

19 Siegel MJ, Bradley EH, Gallo WT, and Kasl SV. The Effect of Spousal Mental and Physical Health on Husbands' and Wives' Depressive Symptoms, Among Older Adults. Journal of Aging and Health 2004; 16: 398-425.

20 Quirouette C and Gold DP. Spousal Characteristics as Predictors of Well-Being in Older Couples. International Journal of Aging and Human Development 1992; 34: 257-269.

21 Larson R, Richards MH (1994) Divergent realities: The emotional lives of mothers, fathers, and adolescents. New York, Basic Books.

22 Hagnell O and Kreitman N. Mental Illness in Married Pairs in a Total Population. British Journal of Psychiatry 1974; 125: 293-302.