A response to the Brown et al critique.

In their "critical reanalysis" of the Fredrickson et al.[1] report noting associations between eudaimonic well-being and reduced expression of adversity-sensitive genes, Brown et al.[2] argue that the results are "no more than a product of chance" and judge "the chances of a successful reproduction… remote."

Their conclusion is invalid for 2 primary reasons:

1) The Fredrickson et al. results have already been replicated.

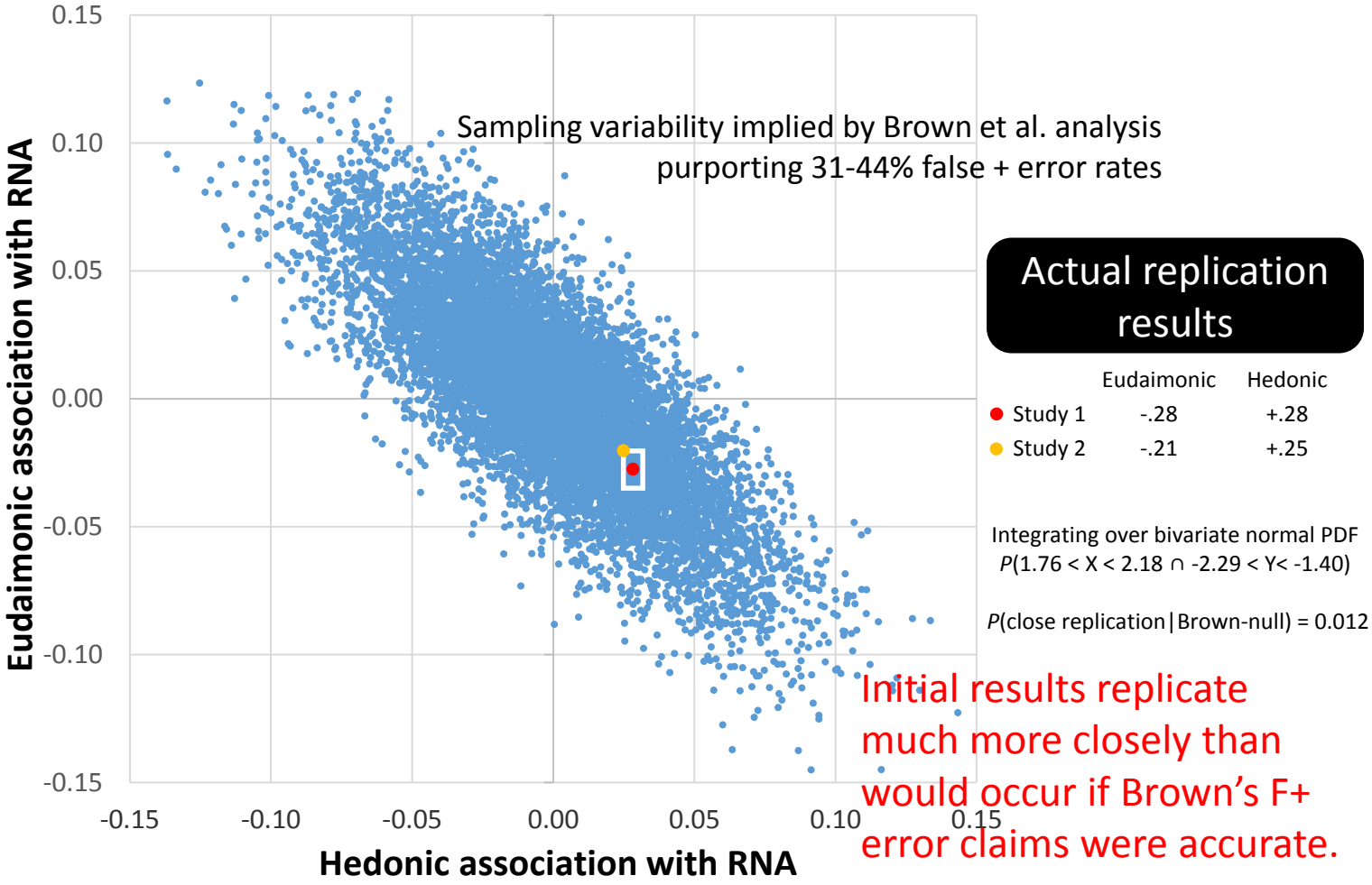
2) The "bitmapping" analysis Brown et al. use to estimate false positive error rates is invalid.

1. Fredrickson et al. (2013) A functional genomic perspective on human well-being. Proc Natl Acad Sci USA, 110(33):13684–13689
2. Brown et al. (2014) A critical reanalysis of the relationship between genomics and well-being. Proc Natl Acad Sci USA, in press. www.pnas.org/cgi/doi/10.1073/pnas.1407057111

1. Replication of the initial result

N = 122 healthy adults

Direct replication of previous measures and analyses, as well as new analyses by mixed effect linear modeling.



2. “Bitmapping” analysis is invalid

Brown et al. conduct systematic combinatorial partitioning of observed psychometric variables and use the results to estimate F+ error rates.

The problem: bitmapping/systematic re-partitioning of a fixed data set is NOT random, does not involve any resampling of observations (subjects), and therefore cannot provide any valid estimate of F+ error rates.

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14
Subject 1	4	4	4	4	3	3	3	3	4	4	5	5	5	4
Subject 2	5	5	5	5	5	3	3	3	5	3	5	5	5	5
Subject 3	4	4	4	4	5	0	4	0	5	4	5	1	4	4
Subject 4	5	5	5	5	5	3	5	5	5	4	5	5	4	4
Subject 5	4	4	4	4	3	1	3	1	3	4	4	4	4	4
Subject 6	3	4	4	4	3	3	4	4	4	4	3	4	4	4
Subject 7	3	4	4	4	5	1	5	5	3	2	5	5	5	2
Subject 8	4	5	4	5	4	2	3	0	4	4	4	4	4	4
Subject 9	5	5	4	4	4	3	3	3	3	3	3	5	4	4
Subject 10	2	2	1	3	2	3	2	1	2	1	1	2	1	1
Subject 11	3	4	3	4	4	1	3	1	3	4	3	3	3	5
Subject 12	3	4	1	4	4	0	4	1	1	2	3	0	4	1
Subject 13	5	5	4	5	5	3	4	2	4	5	5	5	5	5
Subject 14	0	2	1	1	0	0	3	3	0	0	0	0	3	0
Subject 15	5	5	5	5	5	3	5	4	5	3	4	5	5	5

The evidence: striking discontinuity between…

1) true sampling distributions for associations between well-being scores and RNA expression

vs.

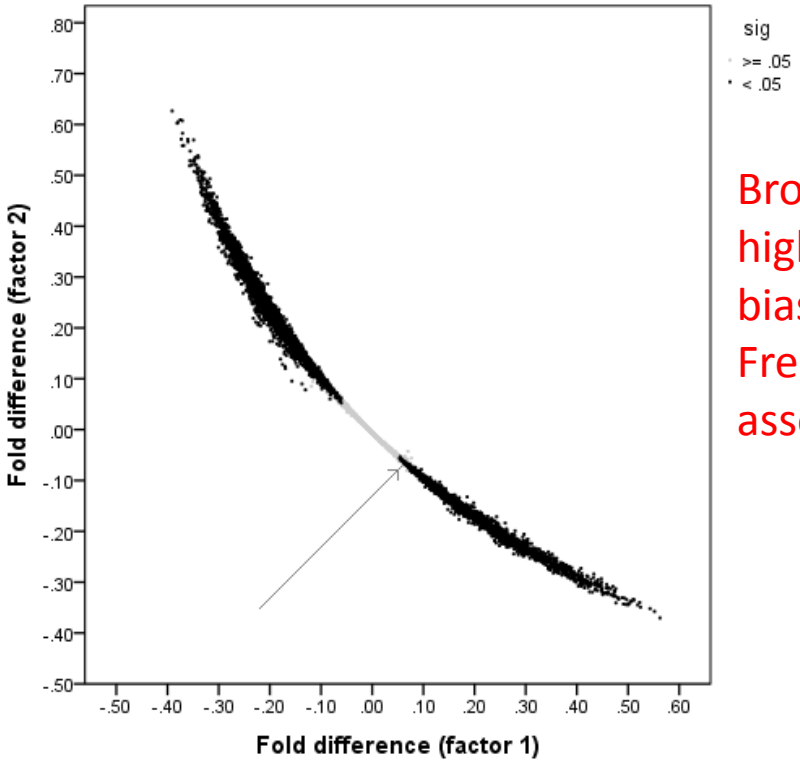
2) distributions emerging from Brown et al.'s "simulation" of Fredrickson et al. data analyses

Parameter estimate

RNA association with well-being scores

“Bitmap” estimate

Figure 7. Scatter plot of 8,191 possible combinations of the 14 items of the MHC-SF into “factors” using psychometric data



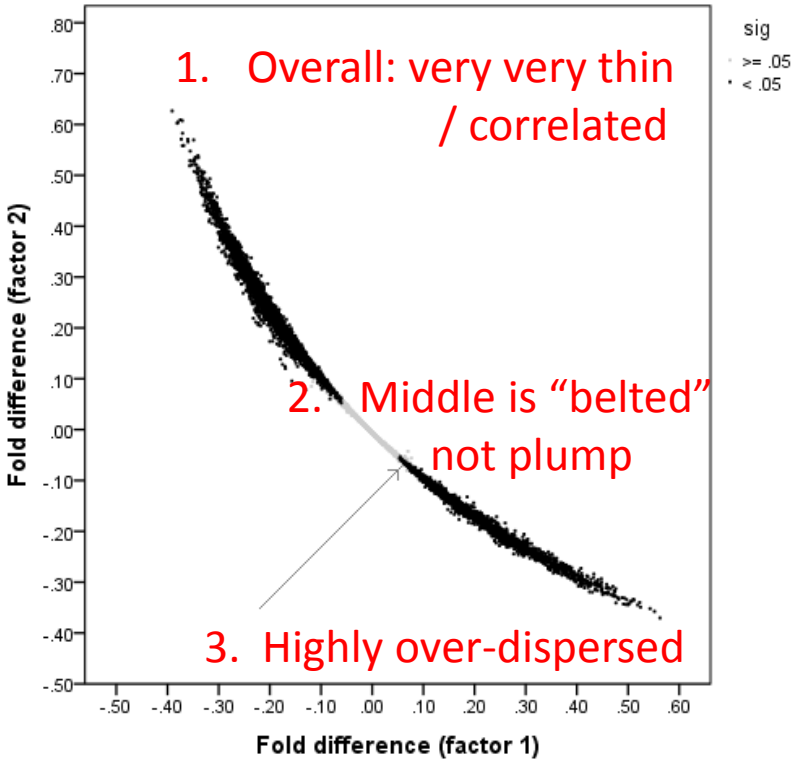
Brown et al. assert the bizarre distribution and high statistical significance rates stem from some bias inherent in the association estimator Fredrickson et al. used to quantify pooled association of 53 indicator mRNAs with well-being.

Parameter estimate

RNA association with well-being scores

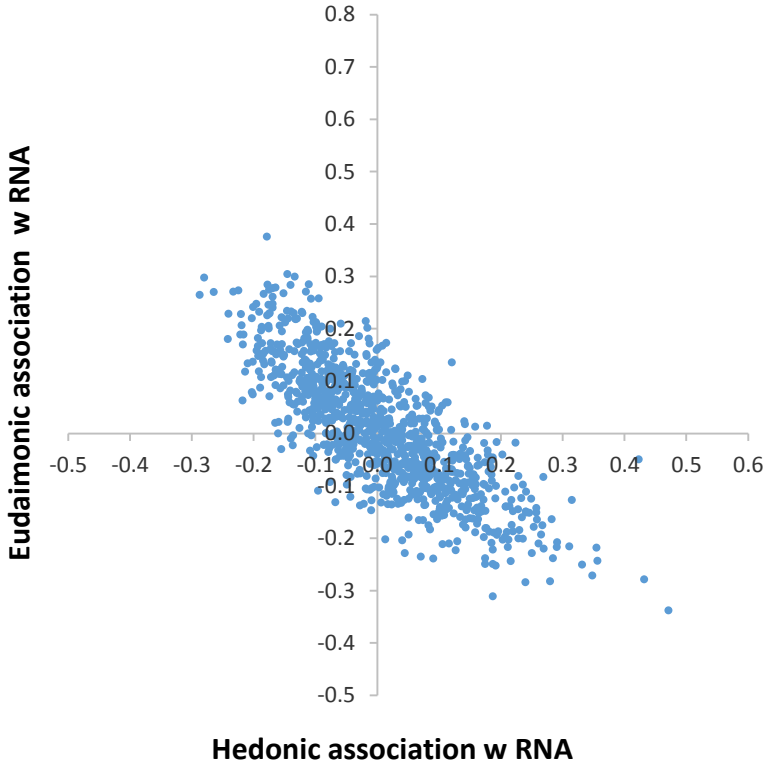
"Bitmap" estimate

Figure 7. Scatter plot of 8,191 possible combinations of the 14 items of the MHC-SF into "factors" using psychometric data



True null sampling distribution

(Permutation)



Bitmapping produces inaccurate estimates of parameter sampling distributions.

Does the problem stem from the data?

One way to tell: feed the bitmap analysis randomly generated data and examine the resulting null distribution for an established benchmark estimator (e.g., 2-sample *t* test)

1. Randomly generate data matrix

   Uniformly distributed integers 0-5, as in Brown's SI Fig 9, centered to mean=0

2. Generate "pseudo factor" scores from bitmap partitions

3. Compute 2-sample *t* test on pseudo factor scores

   Should be completely null, with mean and difference distributions centered on 0

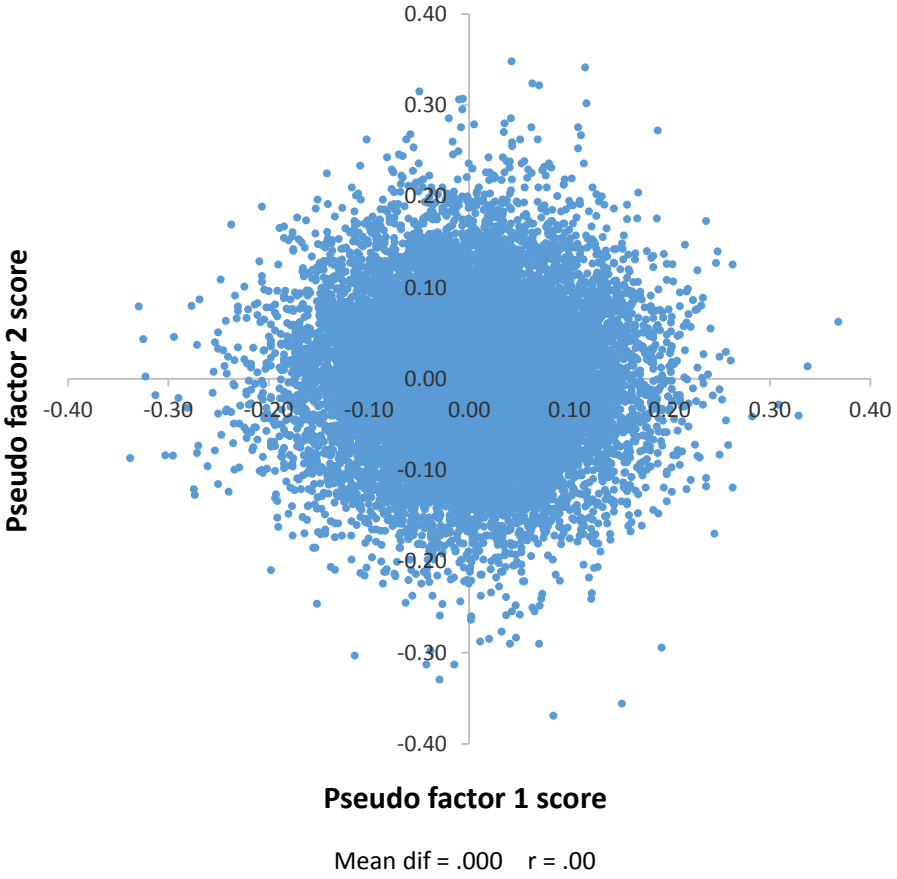
Note: this analysis does not involve

- well-being data
- RNA data
- RNA/well-being association estimator

Parameter estimate

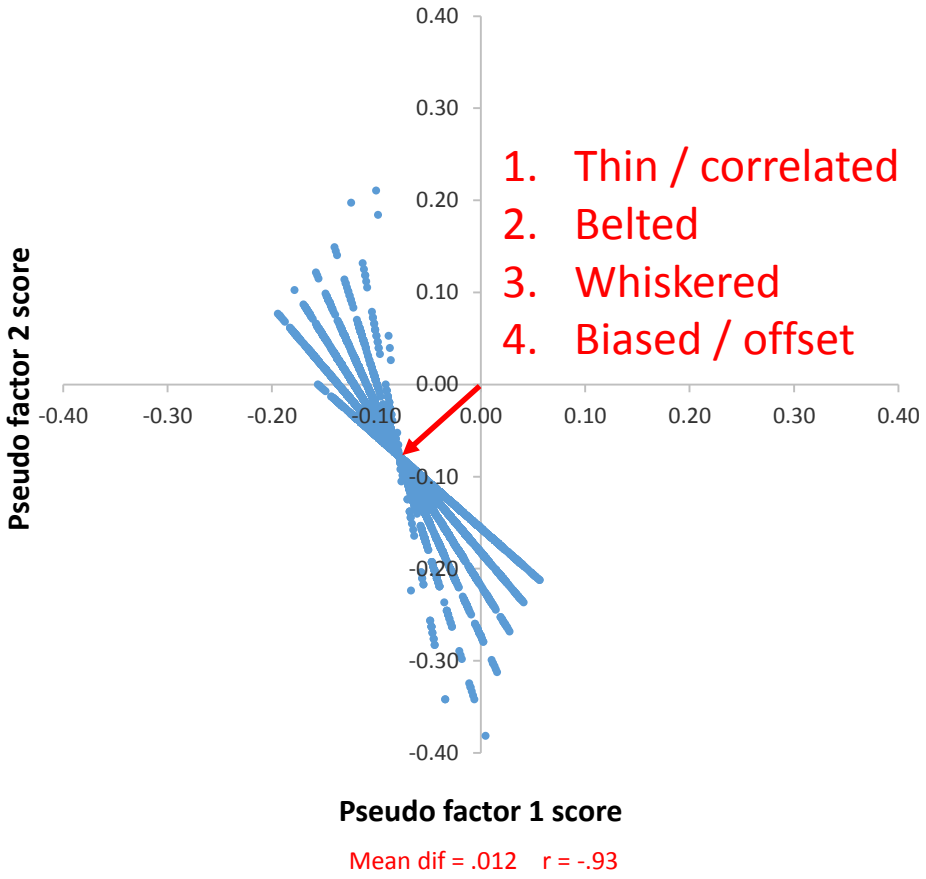
Pseudo factor means

True null sampling distribution



Mean dif = .000 r = .00

Bitmap distribution



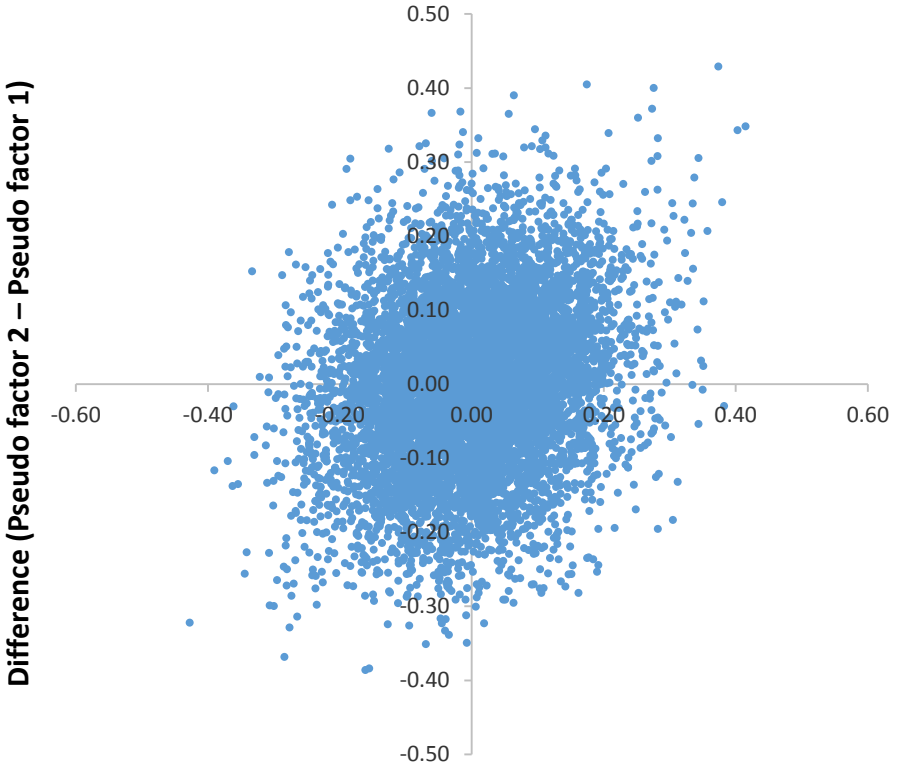
Mean dif = .012 r = -.93

Bitmap procedure generates erroneous sampling distributions for group means.

Parameter estimate

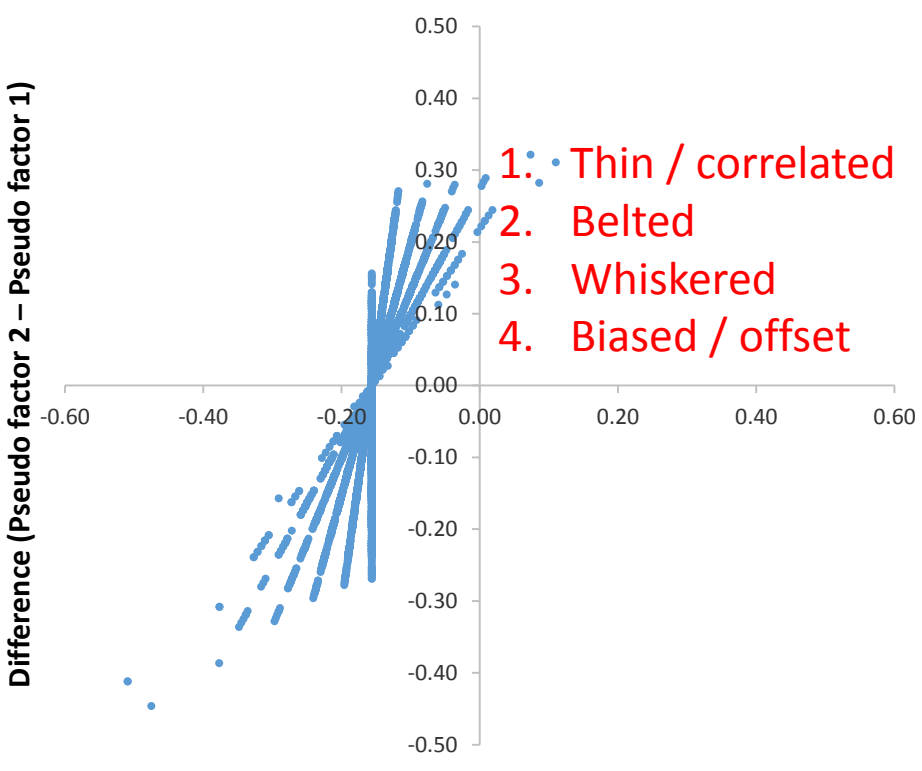
Difference between means

True null sampling distribution



Mean dif = .000

Bitmap distribution



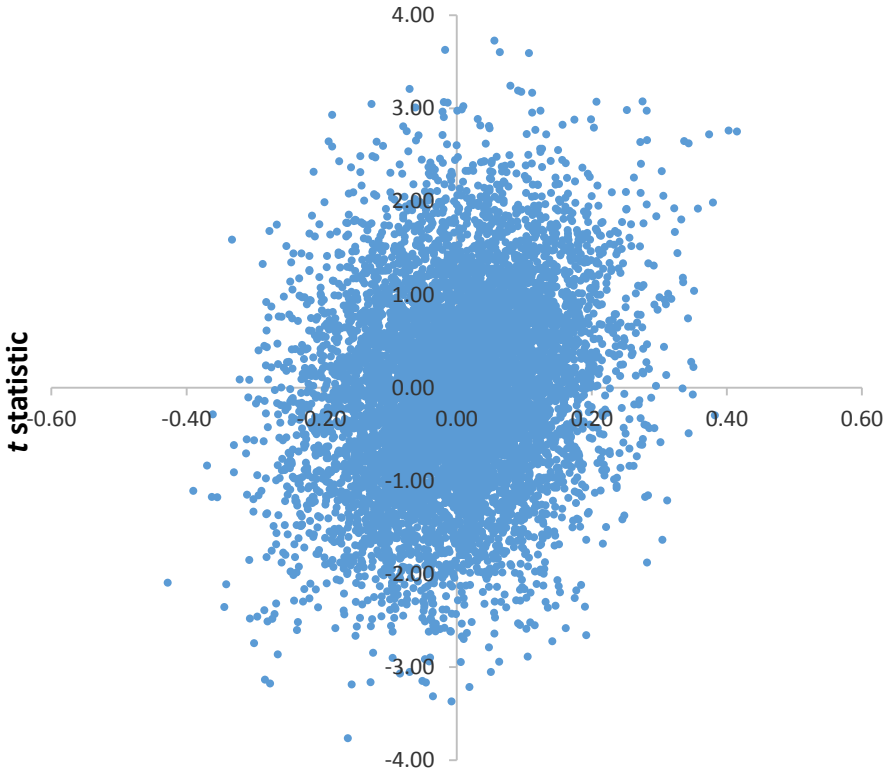
Mean dif = -.012

Bitmap procedure generates erroneous sampling distributions for effect sizes.

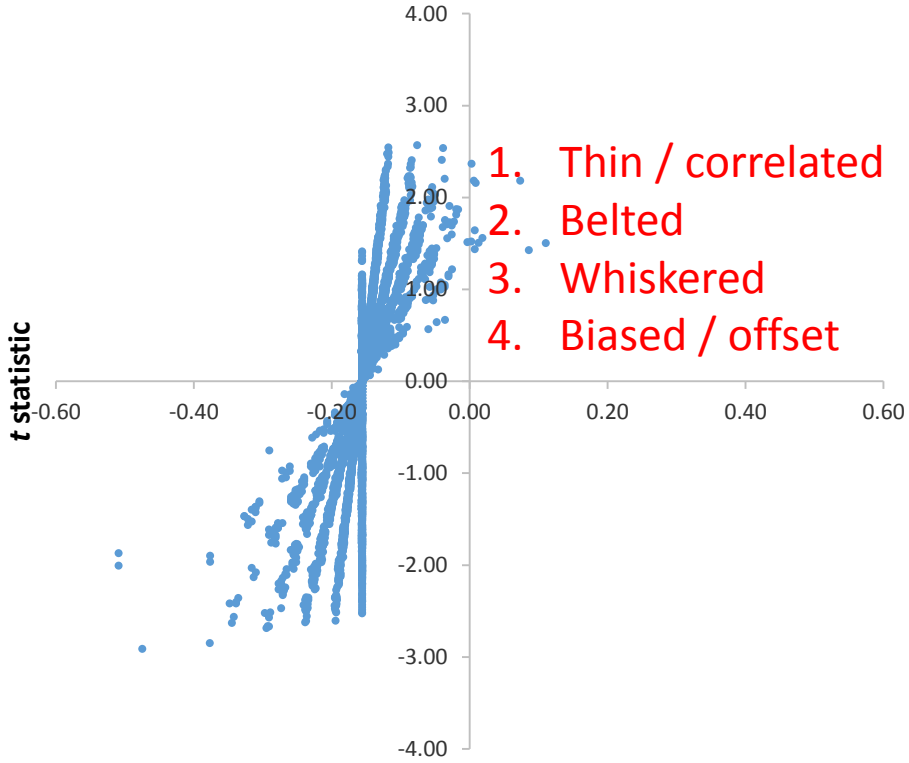
Parameter estimate

2-sample *t* test statistic

True null sampling distribution



Bitmap distribution



Bitmap procedure generates erroneous sampling distributions for test statistics.

Figure 8. Scatter plot of 8,191 MHC-SF “factors”, with psychometric data replaced by normally-distributed random numbers,

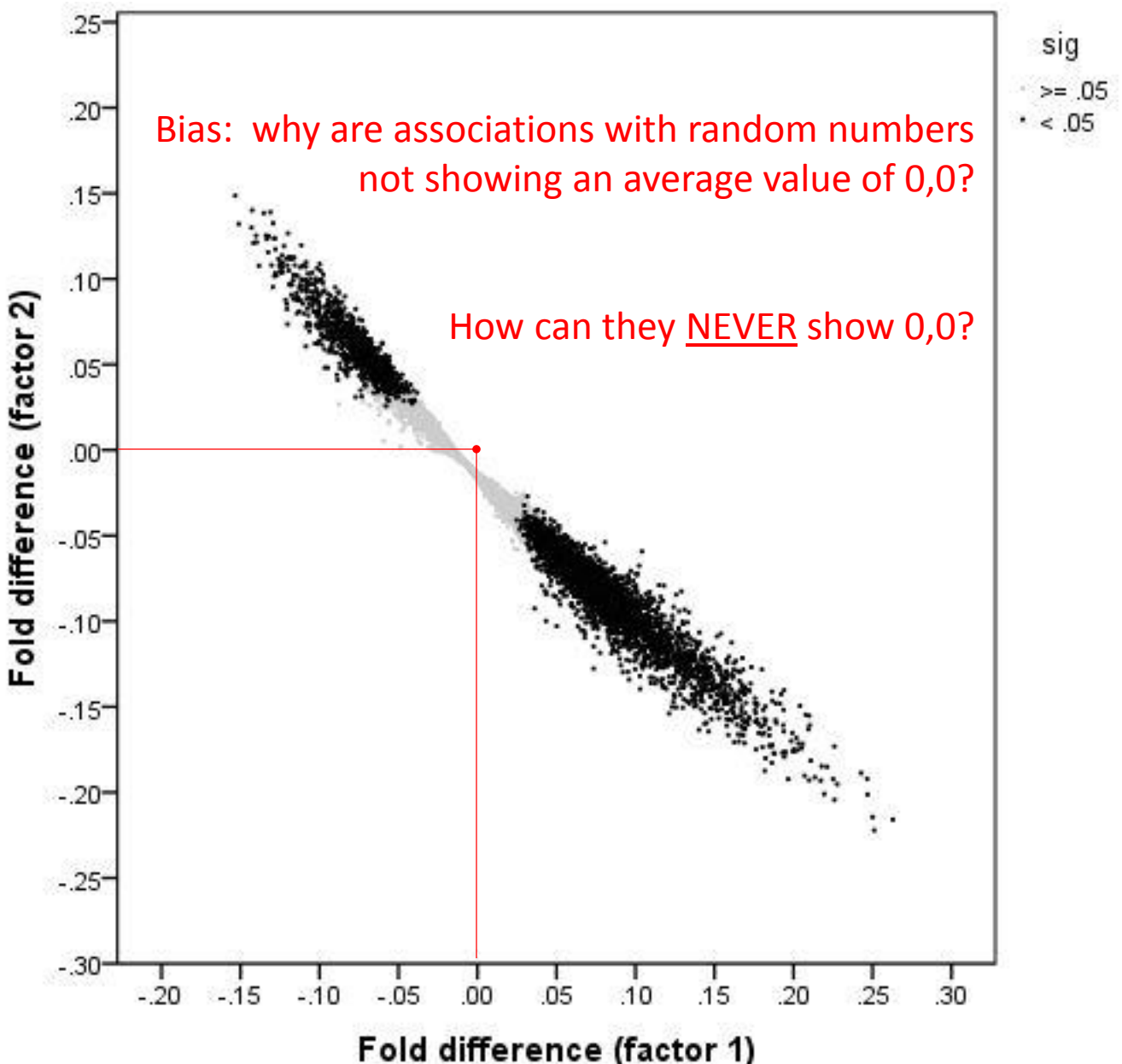
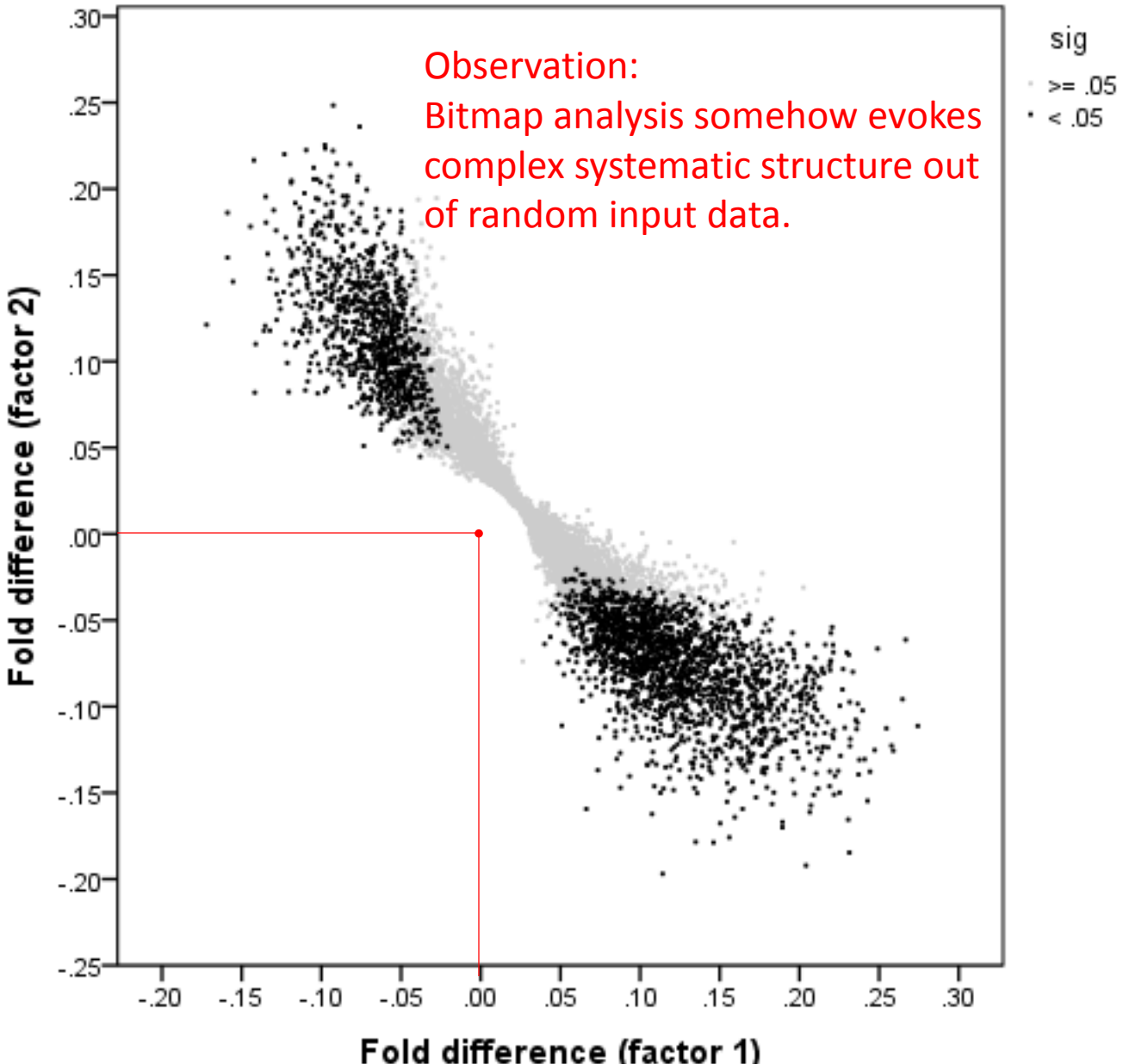


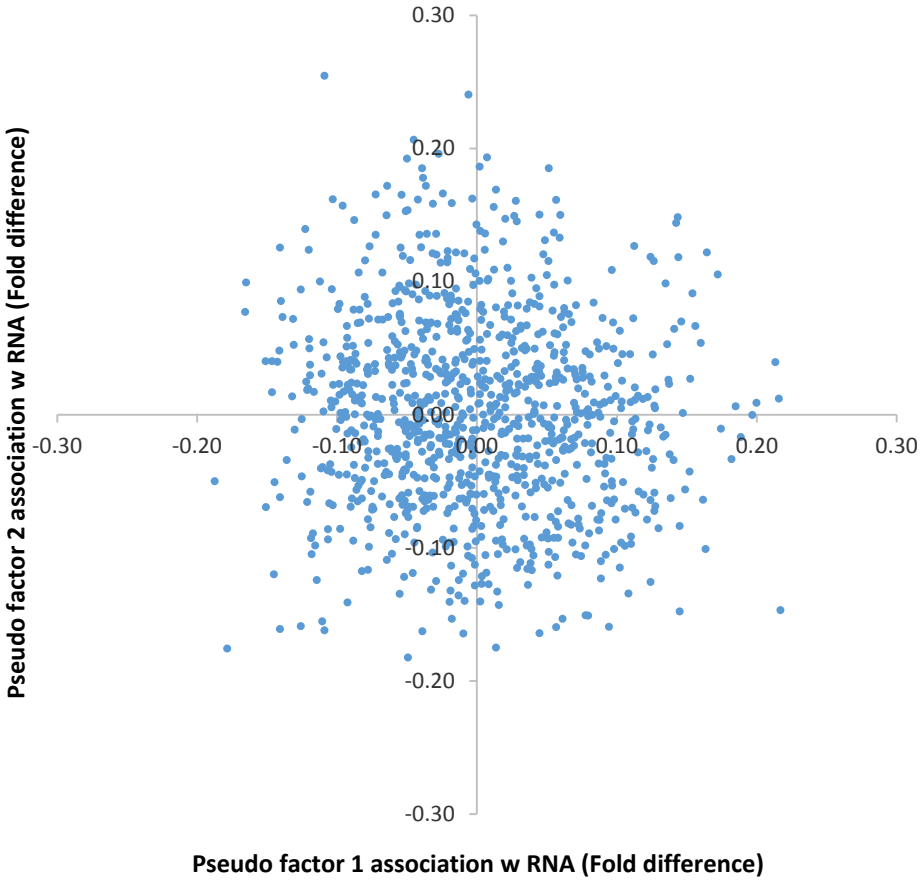
Figure 9. Scatter plot of 8,191 MHC-SF “factors” with psychometric data replaced by uniformly-distributed random numbers



Asymmetry bias

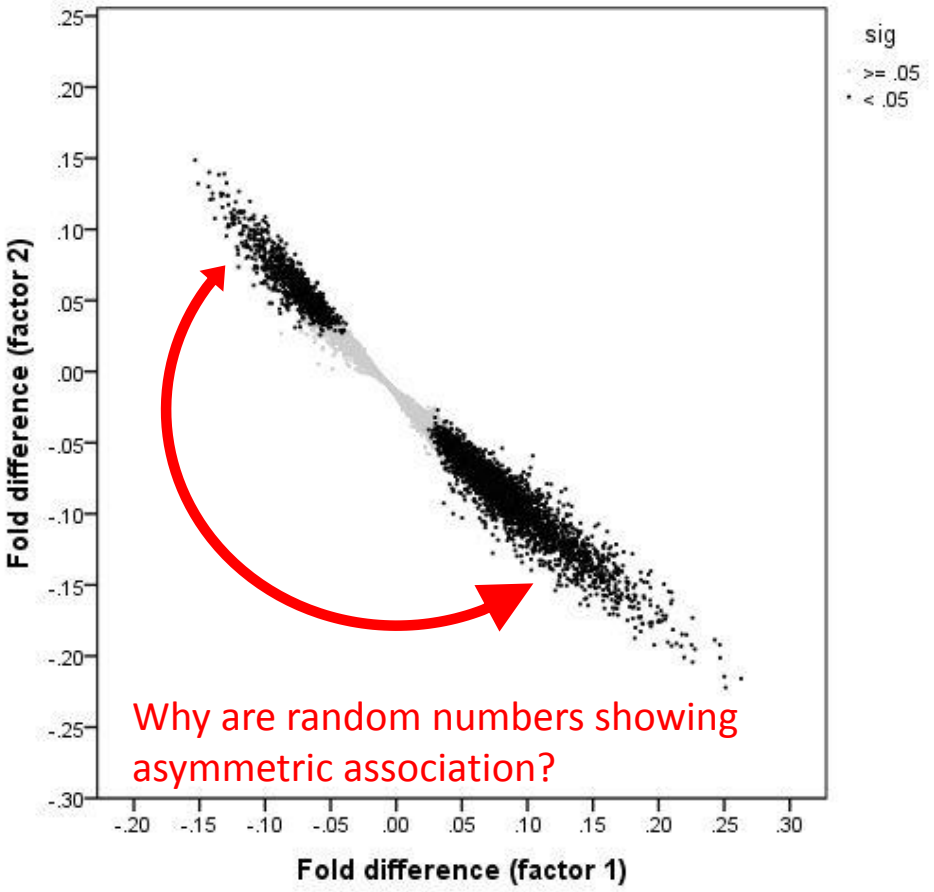
Association of RNA data with randomly distributed normal predictor variables

True null sampling distribution



Bitmap distribution

Figure 8. Scatter plot of 8,191 MHC-SF "factors", with psychometric data replaced by normally-distributed random numbers



OK, the parameter estimates are wrong. But maybe the *p*-values are OK?

First, if the parameter estimates are invalid, then the *p*-values that depend on them are invalid as well.

But just to be sure….

1. Randomly generate data matrix

   Uniformly distributed integers 0-5, as in Brown's SI Fig 9

2. Generate "pseudo factor" scores from bitmap partitions

3. Compute 2-sample *t* test on pseudo factor scores

   p-value distribution should be completely uniform over the range 0-1

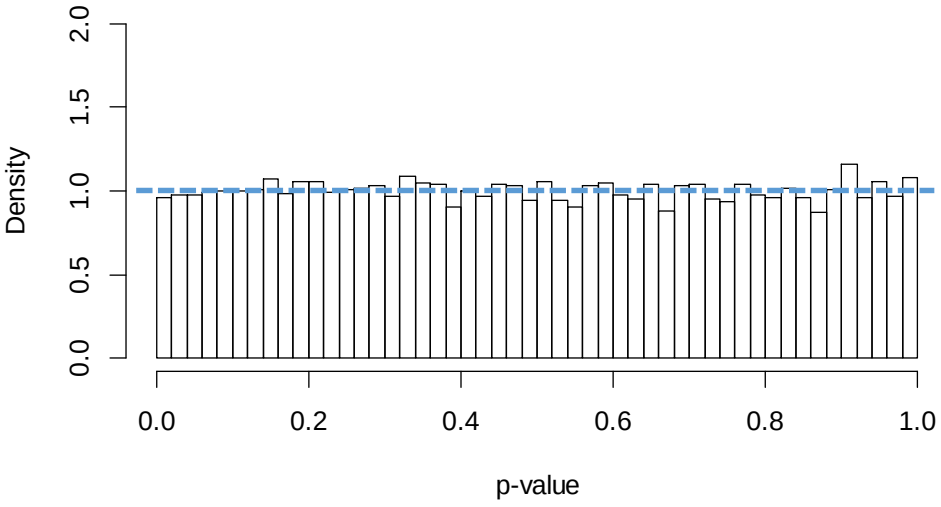
Note: this analysis does not involve

- well-being data
- RNA data
- RNA/well-being association estimator

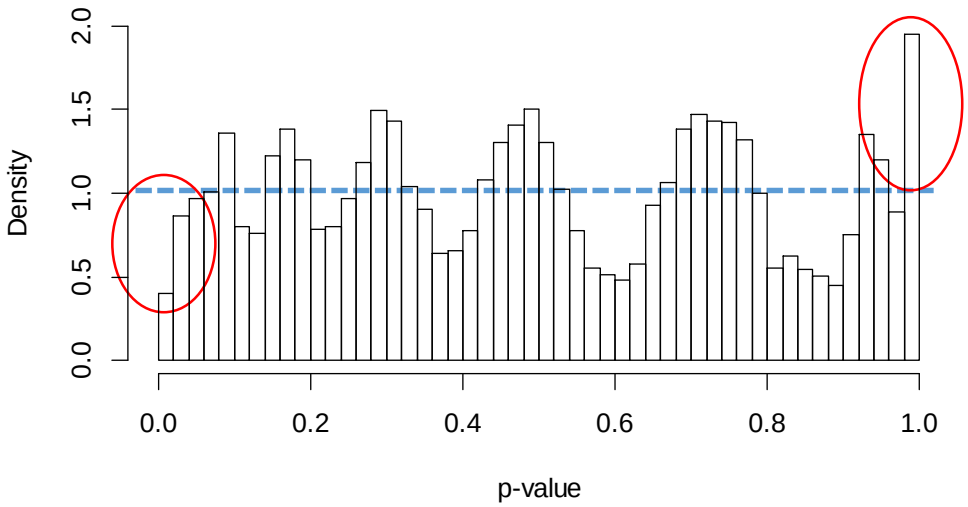
p-value distribution

2-sample *t* test on pseudo factor group means

True null sampling distribution



Bitmap distribution



Bitmapping produces erroneous *p*-value distributions.

Take-home points:

1. Associations between eudiamonic well-being and gene expression have been replicated.

2. The "simulation" results of Brown et al. are invalid and irrelevant.

    - Irrelevant: The bitmapping recombination analysis is not simulating the analysis Fredrickson et al performed.*

    - Invalid: The bitmapping re-partitioning algorithm does not provide a valid assessment of true sampling variability or False + error rates.**

    - Misinterpreted: The aberrant sampling distributions Brown et al. attribute to the RNA association estimator are actually artifacts of their own algorithm.***

* 1. MHC-SF scales were scored according to the developer's established allocation of eudiamonic and hedonic items. Scoring was not based on results of any re-combination of MHC-SF items or any aspect of the observed data.
** 2. Brown et al. offer no citation or mathematical justification for the use of systematic variable repartitioning across a fixed data set to estimate random sampling distributions or F+ error rates. On investigation, the bitmapping method is quickly found to be invalid.
*** 3. Systematic re-partitioning of variables observed on a fixed set of data will produce distorted distributions for any data set analyzed by any statistical procedure (including benchmarks such as the *t* test, and totally random data).

For statisticians: what is "bitmapping?"

Brown et al. take a fixed data set, systematically re-partition it across variables (not subjects), and compute parameter estimates and *p*-values based on data generated by each partition. There is no random sampling of observations at all.

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14
Subject 1	4	4	4	4	3	3	3	3	4	4	5	5	5	4
Subject 2	5	5	5	5	5	3	3	3	5	3	5	5	5	5
Subject 3	4	4	4	4	5	0	4	0	5	4	5	1	4	4
Subject 4	5	5	5	5	5	3	5	5	5	4	5	5	4	4
Subject 5	4	4	4	4	3	1	3	1	3	4	4	4	4	4
Subject 6	3	4	4	4	3	3	4	4	4	4	3	4	4	4
Subject 7	3	4	4	4	5	1	5	5	3	2	5	5	5	2
Subject 8	4	5	4	5	4	2	3	0	4	4	4	4	4	4
Subject 9	5	5	4	4	4	3	3	3	3	3	3	5	4	4

The resulting "sampling" (actually re-combination) distributions yield…

- Biased parameter estimates
- Biased test statistics
- Invalid statistical distributions and *p*-values

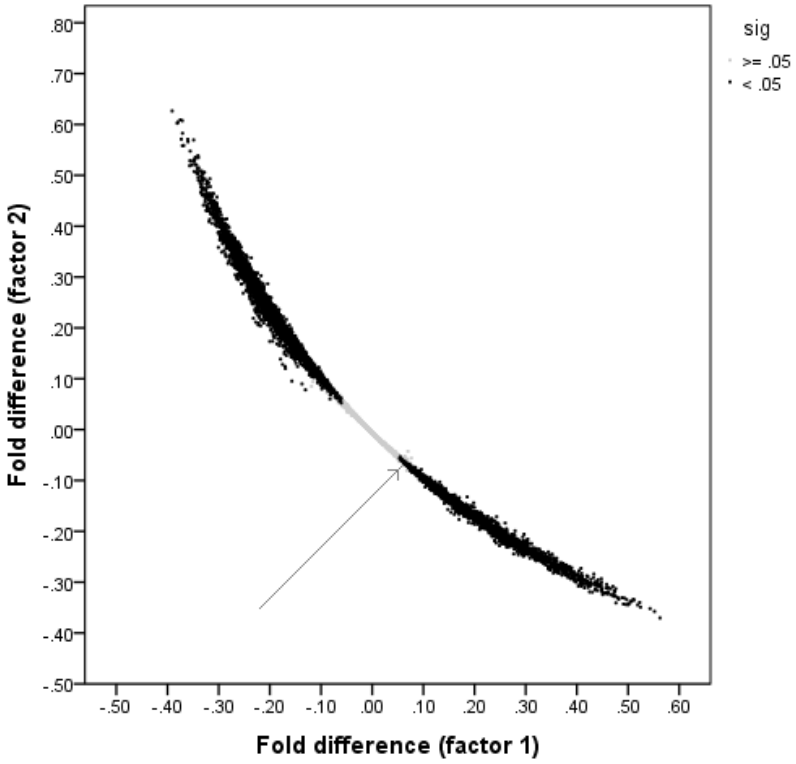
Why? The recombination procedure is fully deterministic, conditional upon a fixed observed data set (either systematic/observed or randomly generated).
Resampling or randomly partitioning across rows/cases is what they should have done.

Computing a distribution across alternative variable partitions (conditional on fixed observations) seems very strange.

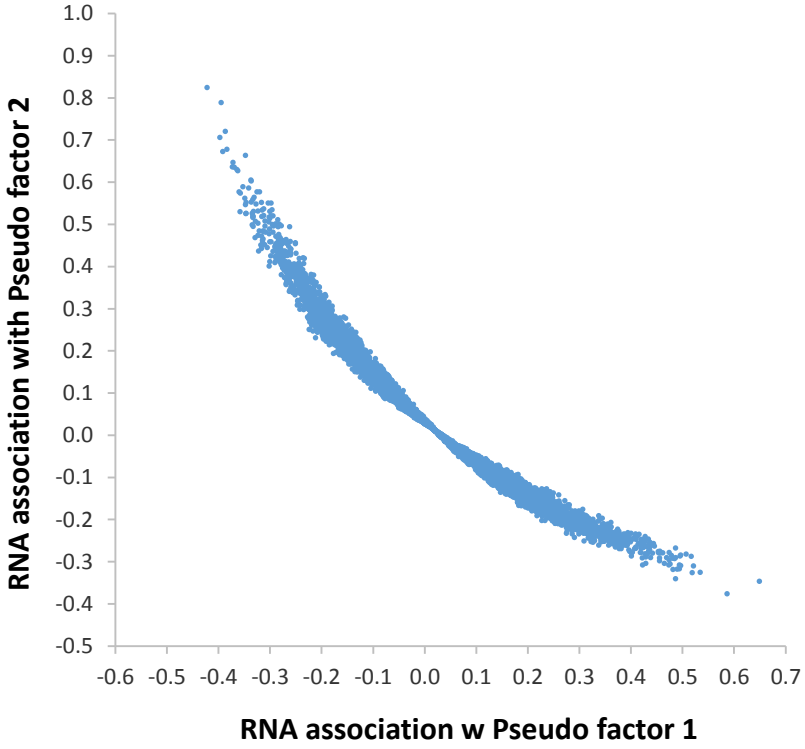
Is that what they really did?

Brown et al. Fig 7

Figure 7. Scatter plot of 8,191 possible combinations of the 14 items of the MHC-SF into "factors" using psychometric data



When we did it.



We understand what they are doing. Regardless of whether it makes sense.

Minor numerical differences are due to rounding error

Why is bitmapping problematic for statistics?

Failure to randomly sample has 2 significant implications:

1) It generally induces non-null statistical distributions (expected value ≠ 0), even from randomly generated input data*

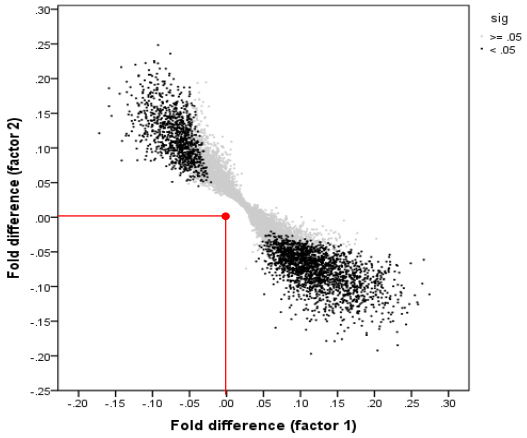
2) It efficiently capitalizes on chance variations in the single fixed data set analyzed, yielding results that are generally unreplicable

* This implies that bitmap distributions cannot generally provide any information about F+ rates (because a true non-zero association is generally present in the bitmap "population")

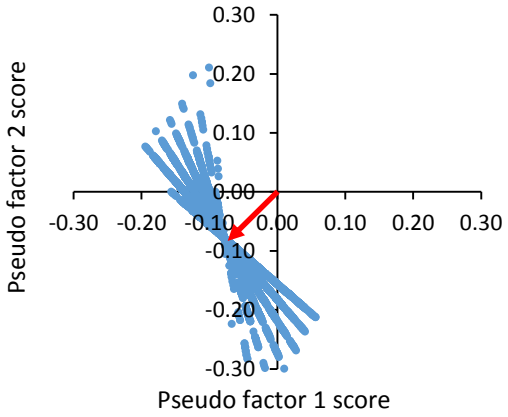
Bitmap-induced bias in statistical distributions

Bitmap-induced bias creates "non-central" distributions that show a TRUE systematic association with outcomes (even when the input data are randomly generated!*).

Brown Figure 9. Scatter plot of 8,191 MHC-SF "factors" with psychometric data replaced by uniformly-distributed random numbers



Randomly generated integers [0,5] centered to true mean = 0 (x - 2.5), and bitmap partitioned into 2 "pseudo-factors"



Implication:

Bitmap distributions cannot provide any valid information about False positive error rates.

Because the distribution shows a true association (expected value ≠ 0,0), statistical significance rates reflect only:

- True positives (significant | true effect)
- False negatives (non-sig | true effect)

p-values provide no information about False positive error rates because F + occurs only in the context of a true null distribution (expected value = 0,0).

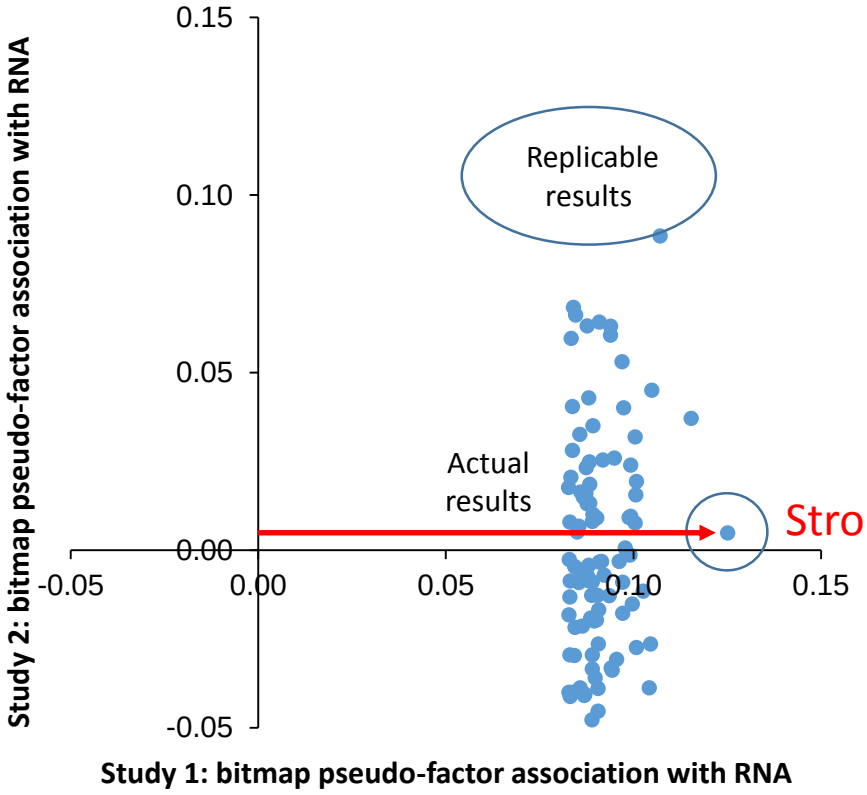
Many/all *p*-values in a bitmap distribution **should** reach statistical significance because the null hypothesis is in fact false (due to bitmap-induced distributional bias). Brown's claim of "inflated significance rates" stems from his own bitmap data manipulation (not from association estimators).

* This stems from the bitmap's systematic repartitioning of a single data set, instead of random resampling of cases.

Why is bitmapping statistically invalid?

Statistics is fundamentally about identifying non-random / replicable associations.
Because bitmapping does not involve any quantification of random sampling variability, it produces fundamentally unreliable / unreplicable findings.

Bitmapping is a system for efficiently capitalizing on chance.



Demonstration:

Test in Study 2 the replicability of Study 1's top 100 bitmap associations with RNA

Study 1: mean = .091 ± .001, $p < 10^{-100}$

Study 2: mean = .001 ± .003, $p = .6971$

Strongest bitmap association in Study 1

Study 1: $z = 16.80$ should be highly replicable

Study 2: $z = 0.16$ no association in Study 2

Test-retest reliability of all 8,191 bitmap associations: $r = .056$

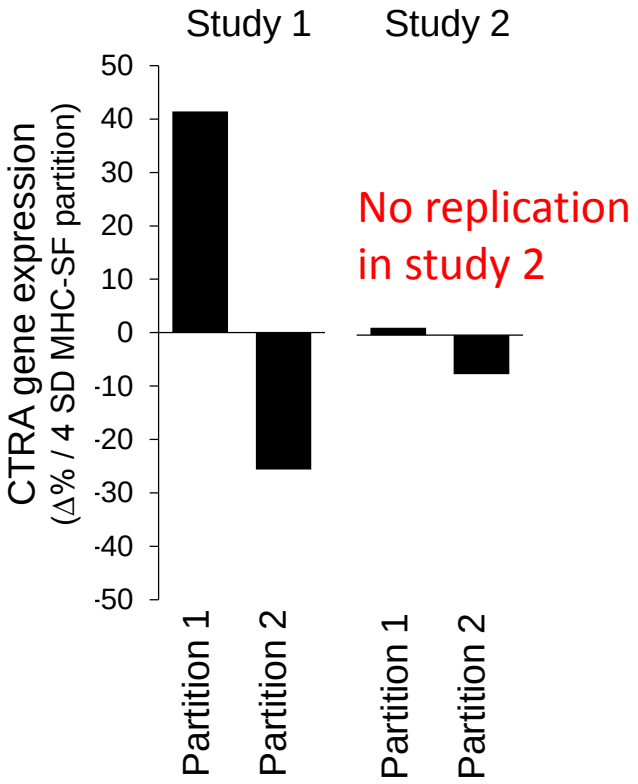
Results from bitmapping analyses are unreliable because bitmapping is statistically invalid.

Why is bitmapping statistically invalid?

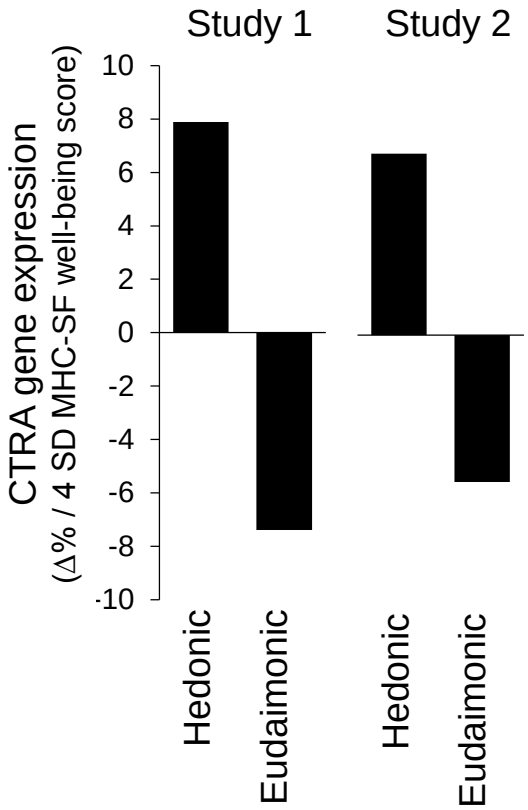
An alternative demonstration based on replicability.

Replicability of Study 1 max bitmap association with RNA

Brown et al.'s "best" pseudo factor pair



Replicability of RNA association with established MHC-SF scores[1]



Bitmapping capitalizes on chance so extremely that is fails to identify reliable findings even when they are present.

1. Scored according to MHC-SF developer's specification of hedonic and eudiamonic items, and used in Fredrickson et al 2013.

For non-statisticians: valid estimation of F+ error rates

False positive error rates for statistical tests are quantified by random sampling simulations, including:

1) **Monte Carlo** analyses, in which random data values are synthesized and fed to the statistical test, and nominally "significant" results are enumerated*

2) **Randomization tests,** in which observed (real) data values are randomly permuted across subjects and fed to the statistical test, and nominally "significant" results are enumerated

3) **Bootstrapping residuals,** in which residuals from observed data are randomly resampled and fed to the statistical test, and nominally "significant" results are enumerated.
(Bootstrapping residuals, rather than observed data values, ensures there is no true association – showing performance under the null hypothesis.)

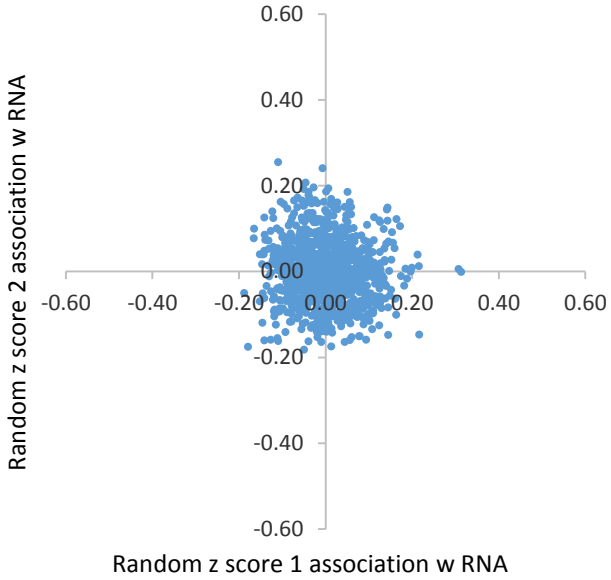
Plots on the following page show results from such analyses

* Supporting Information associated with the original Fredrickson et al. report provided extensive Monte Carlo simulations demonstrating accurate false positive error control for the RNA association estimator. As should be the case – it is simply the sum of random variables, an elementary statistical result.

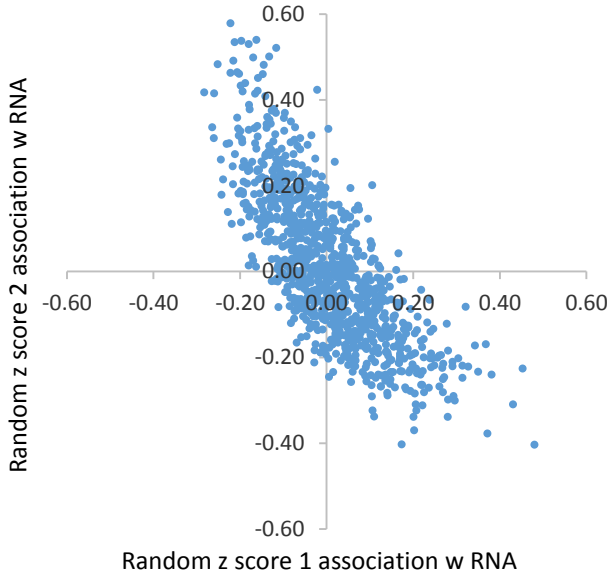
True null hypothesis sampling distributions for the RNA association estimates

(metric = fold-difference parameters graphed by Brown et al.)

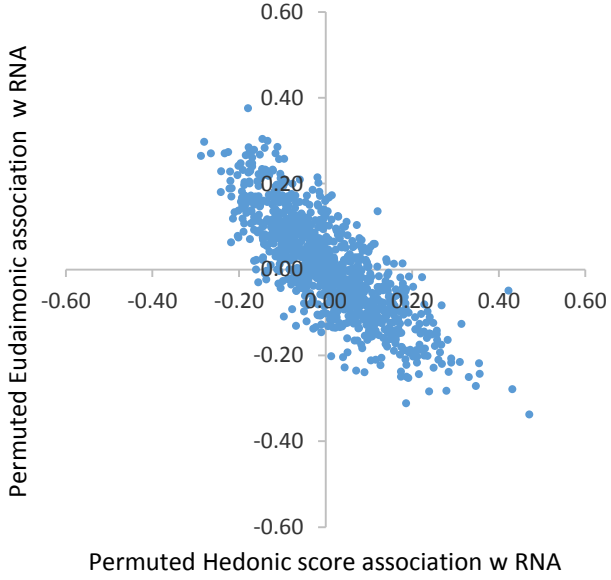
Monte Carlo (2 random *z*, *r* = .00)



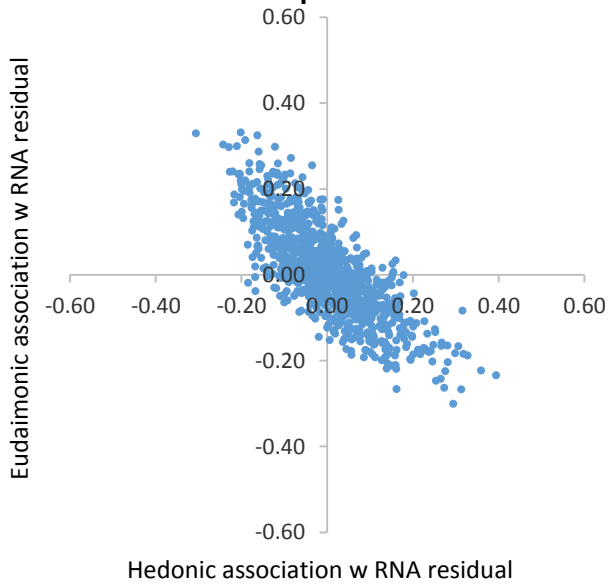
Monte Carlo (2 random *z*, *r* =-.75)



Permutation



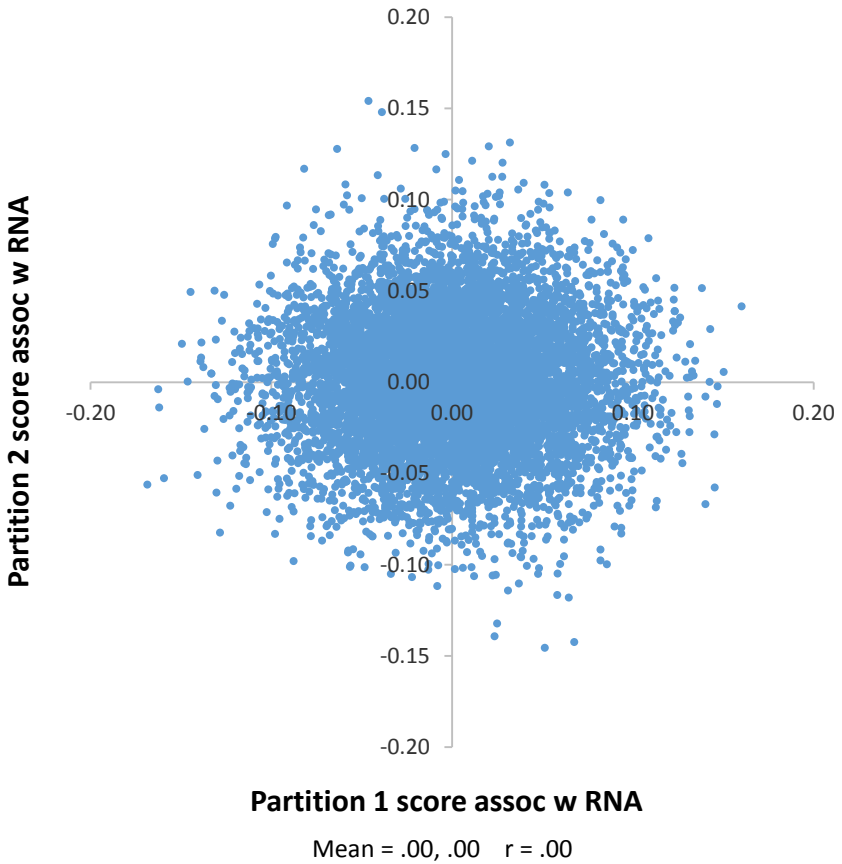
Bootstrap residuals



Parameter estimate

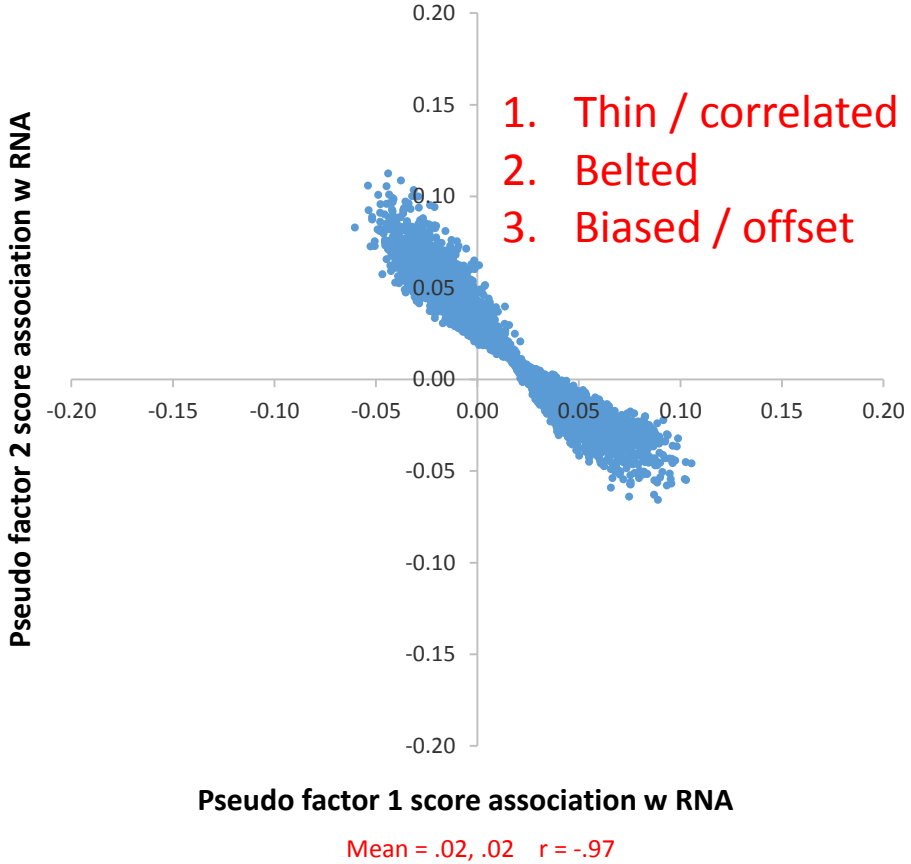
RNA association with random integers – uniform [0,5]

True null sampling distribution



Mean = .00, .00 r = .00

Bitmap distribution



1. Thin / correlated
2. Belted
3. Biased / offset

Mean = .02, .02 r = -.97

Aberrant distribution is created by the bitmap procedure itself.

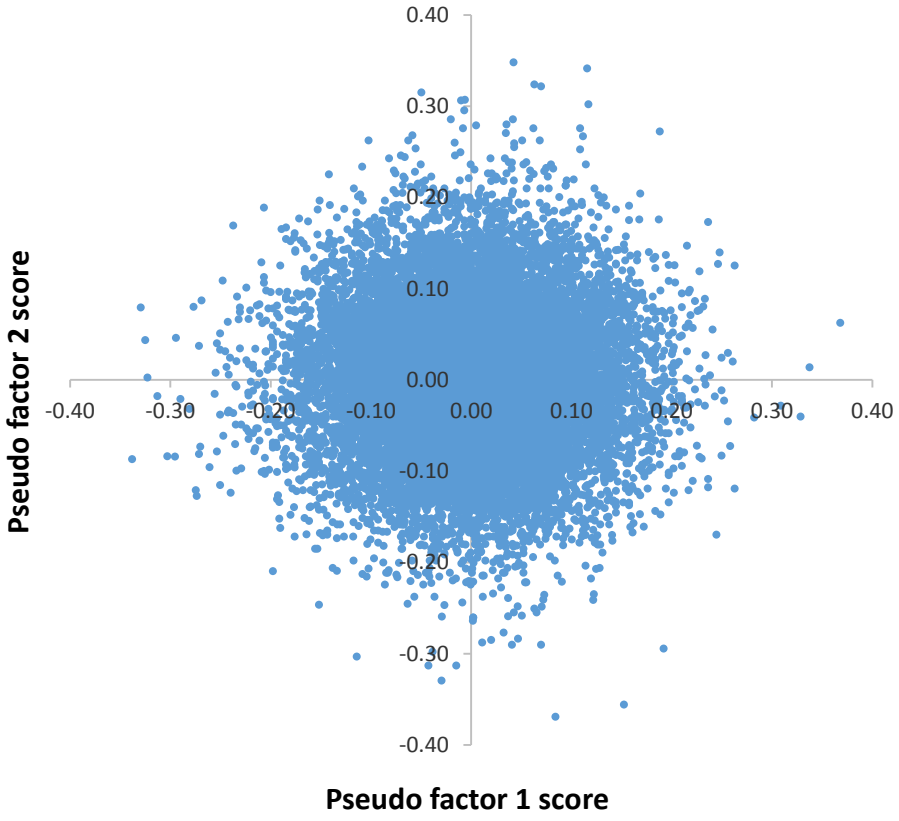
The aberrant distribution is not a product of the RNA pooled association estimate

(which should be obvious, because that is simply an elementary sum of 53 random variables - regression coefficients)

Parameter estimate

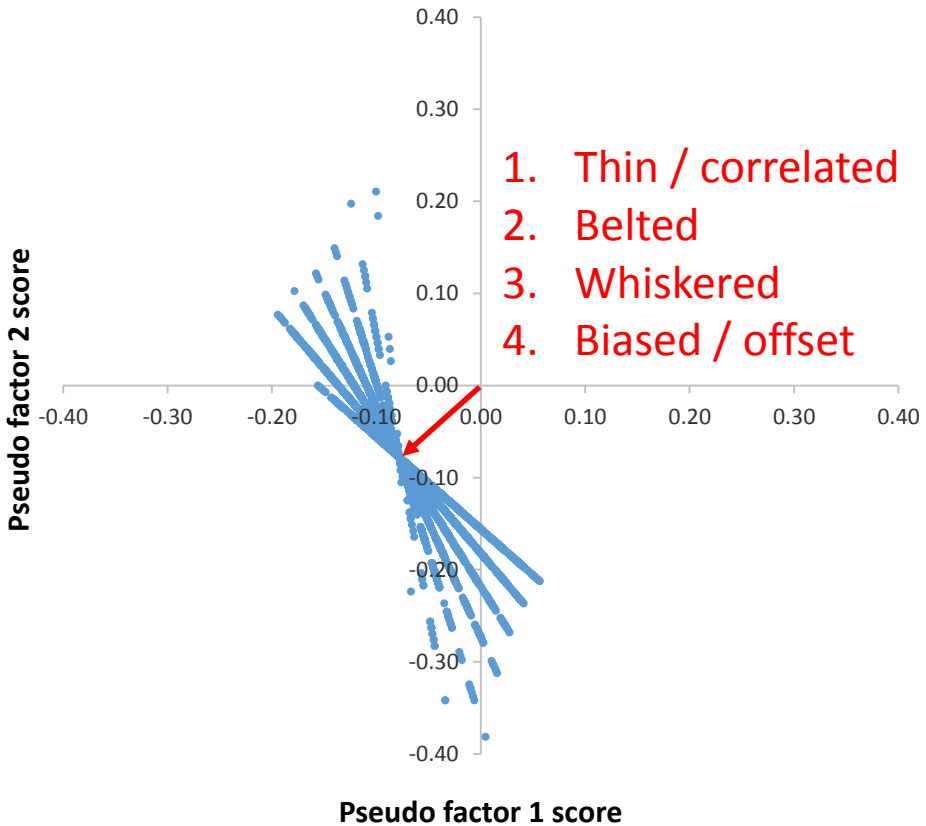
Pseudo factor group means analyzed by 2-sample *t* test

True null sampling distribution



Mean dif = .000 r = .00

Bitmap distribution



Mean dif = .012 r = -.93

Bitmap procedure generates erroneous sampling distributions for group means, estimated differences, test statistics, and *p*-values.

How does such structure arise from "randomness?"

On gloss, bitmapping sounds random because it involves permutation/recombination.
Bitmapping is not as random as it sounds, though, because it involves systematic recombination of variables, rather than the more familiar (and legitimate) random recombination of cases.
Cases are not varied or resampled at all in bitmapping.
As such, bitmapping provides no information about the sampling variability of results, and no information about whether statistically significant results are F+ or T+.

What Brown et al. are actually doing when they claim to test accuracy of the "Fredrickson et al." RNA association analysis:

Data
(observed or random)
→ Bitmap variable re-combinations
8,191 partitions of 14 items
→ Pseudo-factor scores
8,191 pairs of means over partitioned items
→ RNA = f(factor scores)
Predictors = 2 pseudo factor scores reanalyzed 8,191 times

What Fredrickson et al. really did:

Data
(observed or random)
→ 1 established MHC-SF scoring
2 *a priori* scale scores
Based on multiple previous CFAs, Ns ~1000s
No influence of current data on scoring structure.
→ RNA = f(scale scores)
Predictors = 2 *a priori* scale scores

Implications for interpreting Brown et al. results:

Brown et al. draw all of their conclusions regarding False positive error rates from the invalid bitmap analysis. As a consequence:

1. None of the parameter distributions in their SI Figs 7-11 is correct.

2. None of the F+ error rates quoted in the text or the F+ distributions shown in the figures (black vs. grey dots in SI Figs 7-11) is correct.

3. The Brown et al. analysis offers no valid information about likelihood of replication.

Moreover, the result has already been replicated. So, not only is their conclusion analytically wrong, but it is also empirically wrong. The point is moot.