

Information Processing

Pessimism of the Intellect, Optimism of the Will [Favorite posts](#) [Twitter: @hsu_steve](#)



Stephen Hsu
Follow 0

[View my complete profile](#)

Physicist, Startup Founder, Blogger, Dad

Saturday, November 29, 2008

Human genetic variation, Fst and Lewontin's fallacy in pictures

In an earlier post [European genetic substructure](#), I displayed the following graphic, illustrating the genetic clustering of human populations.

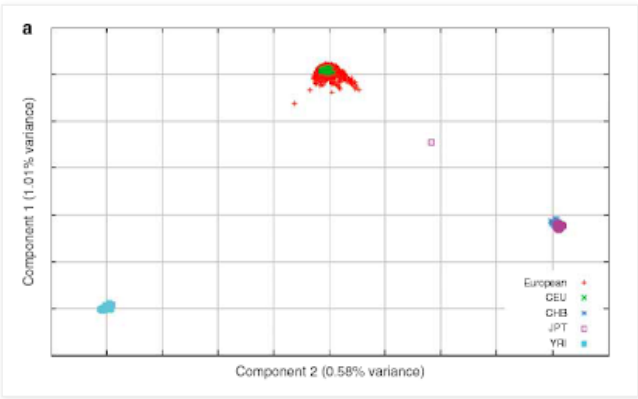


Figure: The three clusters shown above are European (top, green + red), Nigerian (light blue) and E. Asian (purple + blue).

The figure seems to contradict an often stated observation about human genetic diversity, which has become known among experts as [Lewontin's fallacy](#): genetic variation between two random individuals in a given population accounts for 80% or more of the total variation within the entire human population. Therefore, according to the fallacy, any classification of humans into groups ("races") based on genetic information is impossible. ("More variation within groups than between groups.")

To understand this statement better, consider the [F statistic](#) of population genetics, introduced by [Sewall Wright](#):

$$F_{st} = 1 - D_w / D_b$$

D_b and D_w represent the average number of pairwise differences between two individuals sampled from different populations (D_b = "difference between") or the same population (D_w = "difference within"). Even in the most widely separated human populations $F_{st} < .2$ so $D_w / D_b > .8$ (roughly). This may not sound like very much genetic diversity, but it is [more than in many other animal species](#). See [here](#) for recent high statistics F_{st} values by nationality.

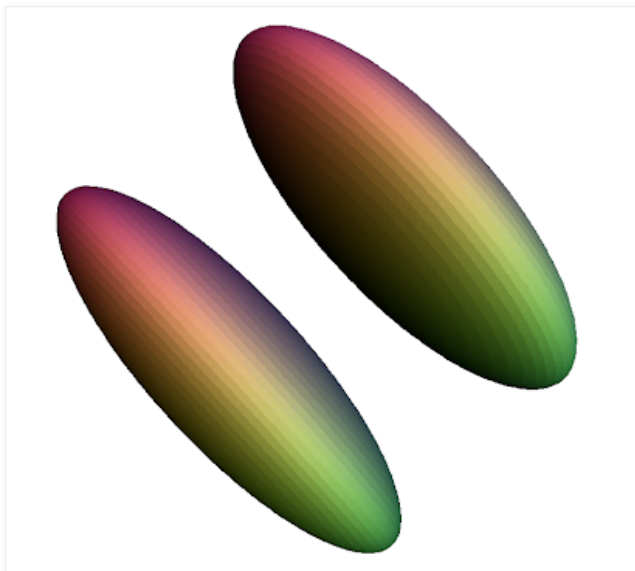
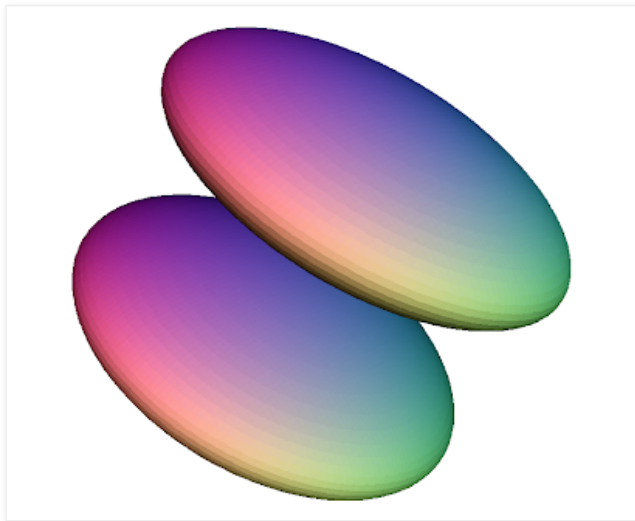
$D_w / D_b > .8$ means that the average genetic distance measured in number of base pair differences between two members of a group (e.g., two randomly selected Europeans) is at least 80 percent of the average distance between distant groups (e.g., Europeans and Asians or Africans). In other words, if two individuals from very distant groups (e.g., a Japanese and a Nigerian) have on average N base pair differences, then two from the same group (e.g., two Nigerians or two Japanese) will on average have roughly $.8 N$ base pair differences.

How can the F_{st} result ("more variation within groups than between groups") be consistent with the clusters shown in the figure? I've had to explain this on numerous occasions, always with great difficulty because the explanation requires a little mathematics. In order to make the point more accessible, I've created the figures below, which show two population clusters,

each represented by an ellipsoid (blob). The different figures depict the same pair of objects, just viewed from different angles.

The blobs are constructed and arranged so that the average distance between two points (individuals) within the same cluster is almost as big as the average distance between two points (individuals) in different clusters. This is easy to achieve if the ellipsoids are big and flat (like pancakes) and placed close to each other along the flat directions. The figure is meant to show how one can have small F_{st} , as in humans, yet easily resolved clusters. The direction in which the gap between the clusters appears is one of the [principal components](#) in the space of human genetic variation, as recently found by bioinformaticists. The figure at the top of this post plots individuals as points in the space generated by the two largest principal components extracted from the combination of data from [HapMap](#) and from [large statistics sampling of Europeans](#). Exhibited this way, isolated clusters ("races") are readily apparent.

The real space of genetic variation has many more than 3 dimensions, so it can't be easily visualized. But some aspects of the figures below still apply: there will be particular directions of variation over which different populations are more or less identical (orthogonal to the principal component; i.e. along the flat directions of each pancake), and there will be directions in which different populations differ radically and have little or no overlap. Note, however, that we are specifically referring to genetic variation, which may or may not translate into phenotypic variation.



Related posts: "no scientific basis for race" , [metric on the space of genomes](#).

The existence of this clustering has been known for [40 years](#).

8 Comments **information processing** Curt Doolittle ▾ Recommend 2 Tweet Share

Sort by Best ▾



Join the discussion...

**Peter Frost** · 10 years ago

Steve,

This all seems to come down to a simple principle of logic: the more you know about something, the more unique it will appear to be. If all I know about a man is his height, it will be impossible to distinguish him from millions of other people. The more pieces of information ('dimensions') I have on him, the less he will overlap with other people. Eventually, with enough information, this man will have zero overlap with others. He will be unique.

Lewontin looked at human genetic variation in terms of one variable at a time. So, perhaps unsurprisingly, he found a lot of overlap.

5 ^ | ▾ · Reply · Share ▾

**Steve Hsu** · 10 years ago

"Eventually, with enough information, this man will have zero overlap with others. He will be unique."

But what is interesting is that long before each individual becomes unique, one can discern discrete clusters that correspond to traditional folk notions of ethnicity.

If you go back and read what Lewontin wrote, he was trying to use his 85-15 statistic to suggest that this would not be the case.

It *could* have been the case that each population overlaps strongly in *each* direction of gene space, which is what Lewontin wanted to suggest. However, it turns out not to be the case...

"...in 1972 Richard Lewontin of Harvard University "found that nearly 85 per cent of humanity's genetic diversity occurs among individuals within a single population." "In other words, two individuals are different because they are individuals, not because they belong to different races." In 2001, the Human Genome edition of Nature(3) came with a compact disc containing a similar statement, quoted above."

<http://www.gnXP.com/MT2/arc...>

3 ^ | ▾ · Reply · Share ▾

**DevilDocNowCiv** → **Steve Hsu** · 4 years ago

Steve,

To cut to the chase with the non-math-savvy among us: Some insist that pesky "Race-IQ Data" should be made to go away because-Gosh! We just noticed! Purely co-incidentally, this fact becomes useful for this pesky "Race-IQ Data"-races don't exist! No need for desegregation-its just confusion! Racism? Just confusion! Go now amongst the masses, ye converted, and spread the good news-they are confused! What they think are different races, some very strongly identifiable, some seemingly an identifiable "mix," and other variants-this is mistaken! Go! Let the vast majority of the world population know it's wrong!

If this idea were truly held, and not promulgated as a "lame" attempt to mitigate the "equality confusion" potentially generated by that pesky "Race-IQ Data," those who argue for the "No Races Theory" would be earnestly spreading the word, and not very, very specifically only using it to mitigate pesky "Race-IQ Data" in article comment blocs and opinion pieces.

1 ^ | ▾ · Reply · Share ▾



Steve Hsu · 10 years ago

I'm not familiar with the 1984 reference, but the modified FST definition used by the human genome project is given here:

<http://en.wikipedia.org/wik...>

"...in the above equation x_{ij} is the estimated frequency (proportion) of the minor allele at SNP i in population j , n_{ij} is the number of genotyped chromosomes at that position, and n_j is the number of chromosomes analysed in that population. The lack of the j subscript in the denominator indicates that statistics n_i and x_i are calculated across the combined data sets."

^ | v · Reply · Share ›



botti → **Steve Hsu** · 6 years ago

***This may not sound like very much genetic diversity, but it is more than in many other animal species. ***

The Goodrum link to the FAQ page doesn't seem to work anymore, but there is a pdf by Goodrum.

<http://stormchan.org/study/...>

Same data also used by

Woodley, 2009. Is Homo sapiens polytypic? Human taxonomic diversity and its implications

<https://lesacreduprintemps1...>

1 ^ | v · Reply · Share ›



Anonymous · 10 years ago

Hi, I was wondering if someone could help me out here. I'm trying to use FST statistics to analyze SNP data from the HGDP database.

But I only seem to find the old version of FST, first introduced by Wright. It seems this formula had some flaws, and now geneticists use the FST formula introduced by Cockerham and Weir (1984). I can't seem to find the formula online, much less an explanation of it.

Anyone have a link to a webpage that deals with the FST formula by Cockerham and Weir?

argiedude

^ | v · Reply · Share ›



Steve Hsu · 10 years ago

The specific data in the first figure is SNPs, although clustering is observed on essentially every type of genetic information thus far examined. I haven't seen specific results for copy number variation -- if you can find some, please let me know.

^ | v · Reply · Share ›



Carson C. Chow · 10 years ago

Nice post Steve. Are these just for SNP's or do they include copy number variations as well?

^ | v · Reply · Share ›

ALSO ON INFORMATION PROCESSING

Information Processing: IceCube: neutrino astronomy in Antarctica

5 comments · 2 months ago

LondonYoung — Tks! That would be pretty decent pay for a grad student! And lots of undisturbed time to study!

Information Processing: Music and Mathematics: Noam Elkies

4 comments · 2 months ago

dxie48 — Re previous: CCR5
delta32https://www.nature.com/arti...Brain enhancement?Some studies have shown that ...

Information Processing: 2019 Digital Business Projections

0 comments · 22 days ago

Information Processing: Manifold Show, episode 2: Bobby Kasthuri and Brain Mapping

12 comments · 6 days ago

9 comments • 22 days ago

Ba'adur Subutai — Stock price will keep pumping as long as QE and zero interest rates co-exist. Almost certainly, the trend will continue in 2019. ...

13 comments • 6 days ago

steve hsu — Thanks -- this is something I've worried about and I'm probably due for a bigger car now that the kids are older.What is the best ...

Subscribe Add Disqus to your site Add Disqus Add Disqus! Privacy Policy Privacy Policy Privacy Policy

Newer Post Home Older Post

Subscribe to: Post Comments (Atom)



Stephen Hsu

Follow 0

View my complete profile

Blog Archive

- ▶ 2019 (12)
- ▶ 2018 (128)
- ▶ 2017 (150)
- ▶ 2016 (206)
- ▶ 2015 (167)
- ▶ 2014 (184)
- ▶ 2013 (211)
- ▶ 2012 (222)
- ▶ 2011 (266)
- ▶ 2010 (234)
- ▶ 2009 (204)
- ▼ 2008 (255)
 - ▶ 12 (25)
 - ▼ 11 (31)
 - Human genetic variation, Fst and Lewontin's fallac...
 - The best and the brightest: McGeorge Bundy
 - Atlas Shrugged, updated
 - Back in the day
 - East Asian genetic substructure
 - The value of trust
 - European genetic substructure
 - Physics, complex systems and economics
 - Perimeter photos
 - Deflation
 - Bill Janeway interview
 - Kakutani on Gladwell
 - Perimeter talk: monsters
 - Central limit theorem and securitization: how to b...
 - More Soros
 - IQ and longevity
 - Bay area housing market has cracked
 - Venn diagram for economics
 - Money men congressional testimony
 - Quants, speak!
 - Michael Lewis on the subprime bubble
 - Fear and loathing of the plutocracy
 - AIG watch
 - Wealth effect, consumer spending and recession
 - Catastrophe bonds and the investor's choice proble...
 - OBAMA!

[Gary Gorton on YSM panel](#)
[AIG killed by CDS collateral calls](#)
[Catastrophe bonds](#)
[My talk on the financial crisis](#)
[The heterodoxy strikes back](#)

- ▶ [10](#) (34)
- ▶ [09](#) (31)
- ▶ [08](#) (16)
- ▶ [07](#) (16)
- ▶ [06](#) (19)
- ▶ [05](#) (21)
- ▶ [04](#) (16)
- ▶ [03](#) (20)
- ▶ [02](#) (10)
- ▶ [01](#) (16)
- ▶ [2007](#) (196)
- ▶ [2006](#) (154)
- ▶ [2005](#) (253)
- ▶ [2004](#) (100)

Labels

- [physics](#) (347)
- [genetics](#) (281)
- [finance](#) (261)
- [globalization](#) (258)
- [brainpower](#) (248)
- [technology](#) (205)
- [genomics](#) (199)
- [american society](#) (194)
- [China](#) (180)
- [economics](#) (168)
- [psychometrics](#) (166)
- [innovation](#) (161)
- [photos](#) (160)
- [science](#) (159)
- [psychology](#) (147)
- [ai](#) (141)
- [travel](#) (136)
- [universities](#) (119)
- [machine learning](#) (116)
- [credit crisis](#) (115)
- [higher education](#) (114)
- [biology](#) (111)
- [human capital](#) (110)
- [iq](#) (103)
- [genetic engineering](#) (102)
- [startups](#) (100)
- [cognitive science](#) (97)
- [careers](#) (80)
- [autobiographical](#) (78)
- [credit crunch](#) (78)
- [statistics](#) (75)
- [elitism](#) (72)
- [evolution](#) (71)
- [gilded age](#) (71)
- [income inequality](#) (69)
- [quantum mechanics](#) (67)
- [political correctness](#) (65)
- [genius](#) (62)

- [politics](#) (62)
- [talks](#) (62)
- [books](#) (61)
- [caltech](#) (60)
- [social science](#) (60)
- [geopolitics](#) (59)
- [mma](#) (55)
- [history of science](#) (53)
- [sci fi](#) (51)
- [academia](#) (50)
- [bounded rationality](#) (50)
- [harvard](#) (50)
- [kids](#) (50)
- [silicon valley](#) (50)
- [podcasts](#) (49)
- [bgi](#) (48)
- [education](#) (48)
- [mathematics](#) (47)
- [MSU](#) (46)
- [cdo](#) (45)
- [realpolitik](#) (44)
- [derivatives](#) (42)
- [behavioral economics](#) (39)
- [jiujitsu](#) (39)
- [history](#) (38)
- [intellectual history](#) (38)
- [literature](#) (38)
- [neuroscience](#) (37)
- [physical training](#) (37)
- [bubbles](#) (36)
- [mortgages](#) (36)
- [ufc](#) (36)
- [computing](#) (35)
- [film](#) (35)
- [bjj](#) (34)
- [expert prediction](#) (32)
- [google](#) (32)
- [hedge funds](#) (32)
- [foo camp](#) (30)
- [many worlds](#) (30)
- [race relations](#) (30)
- [affirmative action](#) (29)
- [black holes](#) (29)
- [economic history](#) (29)
- [quants](#) (29)
- [video](#) (29)
- [efficient markets](#) (28)
- [movies](#) (28)
- [feynman](#) (27)
- [nuclear weapons](#) (27)
- [sports](#) (27)
- [von Neumann](#) (27)
- [biotech](#) (26)
- [music](#) (26)
- [entrepreneurs](#) (25)
- [housing](#) (25)
- [obama](#) (25)
- [security](#) (25)
- [singularity](#) (25)
- [subprime](#) (25)
- [taiwan](#) (24)
- [berkeley](#) (23)
- [conferences](#) (22)

- [meritocracy](#) (22)
- [ultimate fighting](#) (22)
- [athletics](#) (21)
- [wall street](#) (21)
- [cds](#) (20)
- [internet](#) (20)
- [quantum field theory](#) (20)
- [scifoo](#) (20)
- [blogging](#) (19)
- [venture capital](#) (19)
- [gender](#) (18)
- [goldman sachs](#) (18)
- [treasury bailout](#) (17)
- [university of oregon](#) (17)
- [algorithms](#) (16)
- [cryptography](#) (16)
- [freeman dyson](#) (16)
- [new yorker](#) (16)
- [smpy](#) (16)
- [autism](#) (15)
- [cosmology](#) (15)
- [japan](#) (15)
- [oppenheimer](#) (15)
- [personality](#) (15)
- [happiness](#) (14)
- [height](#) (14)
- [wwii](#) (14)
- [chess](#) (13)
- [christmas](#) (13)
- [fitness](#) (13)
- [hedonic treadmill](#) (13)
- [probability](#) (13)
- [aspergers](#) (12)
- [dna](#) (12)
- [india](#) (12)
- [les grandes ecoles](#) (12)
- [malcolm gladwell](#) (12)
- [neanderthals](#) (12)
- [net worth](#) (12)
- [nobel prize](#) (12)
- [privacy](#) (12)
- [social networks](#) (12)
- [television](#) (12)
- [blade runner](#) (11)
- [entropy](#) (11)
- [geeks](#) (11)
- [nsa](#) (11)
- [philosophy of mind](#) (11)
- [string theory](#) (11)
- [ability](#) (10)
- [football](#) (10)
- [government](#) (10)
- [harvard society of fellows](#) (10)
- [italy](#) (10)
- [mutants](#) (10)
- [nerds](#) (10)
- [olympics](#) (10)
- [pseudoscience](#) (10)
- [Einstein](#) (9)
- [Go](#) (9)
- [complexity](#) (9)
- [crossfit](#) (9)
- [dating](#) (9)

- encryption (9)
- eugene (9)
- flynn effect (9)
- research (9)
- turing test (9)
- alan turing (8)
- art (8)
- data mining (8)
- determinism (8)
- james salter (8)
- keynes (8)
- manhattan (8)
- pca (8)
- philip k. dick (8)
- pop culture (8)
- real estate (8)
- robot genius (8)
- usain bolt (8)
- war (8)
- aig (7)
- alpha (7)
- ashkenazim (7)
- basketball (7)
- energy (7)
- france (7)
- free will (7)
- fx (7)
- game theory (7)
- games (7)
- hugh everett (7)
- poker (7)
- qcd (7)
- russia (7)
- success (7)
- Fermi problems (6)
- anthropic principle (6)
- bayes (6)
- class (6)
- climate change (6)
- democracy (6)
- environmentalism (6)
- nassim taleb (6)
- noam chomsky (6)
- paris (6)
- patents (6)
- prostitution (6)
- quantum computers (6)
- tail risk (6)
- teaching (6)
- volatility (6)
- academia sinica (5)
- bobby fischer (5)
- cold war (5)
- econ talk (5)
- godel (5)
- information theory (5)
- iraq war (5)
- kasparov (5)
- luck (5)
- new york times (5)
- nonlinearity (5)
- perimeter institute (5)
- rationality (5)

- renaissance technologies (5)
- sad but true (5)
- software development (5)
- vietnam war (5)
- warren buffet (5)
- 100m (4)
- Iran (4)
- Poincare (4)
- bill gates (4)
- borges (4)
- cambridge uk (4)
- charles darwin (4)
- creativity (4)
- drones (4)
- fake alpha (4)
- feminism (4)
- global warming (4)
- hormones (4)
- humor (4)
- inequality (4)
- intellectual property (4)
- judo (4)
- kerviel (4)
- markets (4)
- microsoft (4)
- mixed martial arts (4)
- monsters (4)
- moore's law (4)
- solar energy (4)
- soros (4)
- trento (4)
- 200m (3)
- babies (3)
- brain drain (3)
- censorship (3)
- charlie munger (3)
- chet baker (3)
- correlation (3)
- demographics (3)
- ecosystems (3)
- equity risk premium (3)
- facebook (3)
- fannie (3)
- fst (3)
- intellectual ventures (3)
- jim simons (3)
- language (3)
- lee kwan yew (3)
- lewontin fallacy (3)
- lhc (3)
- magic (3)
- michael lewis (3)
- nathan myhrvold (3)
- neal stephenson (3)
- olympiads (3)
- path integrals (3)
- risk preference (3)
- search (3)
- sec (3)
- simulation (3)
- sivs (3)
- society generale (3)
- supercomputers (3)

- thailand (3)
- alibaba (2)
- assortative mating (2)
- bear stearns (2)
- bruce springsteen (2)
- charles babbage (2)
- cheng ting hsu (2)
- cloning (2)
- david mamet (2)
- digital books (2)
- donald mackenzie (2)
- eliot spitzer (2)
- empire (2)
- exchange rates (2)
- frauds (2)
- freddie (2)
- gaussian copula (2)
- heinlein (2)
- industrial revolution (2)
- james watson (2)
- ltcm (2)
- mating (2)
- mba (2)
- mccain (2)
- mit (2)
- monkeys (2)
- national character (2)
- nicholas metropolis (2)
- no holds barred (2)
- offices (2)
- oligarchs (2)
- palin (2)
- population structure (2)
- prisoner's dilemma (2)
- skidelsky (2)
- socgen (2)
- sprints (2)
- systemic risk (2)
- twitter (2)
- ussr (2)
- variance (2)
- virtual reality (2)
- abx (1)
- anathem (1)
- andrew lo (1)
- antikythera mechanism (1)
- athens (1)
- atlas shrugged (1)
- ayn rand (1)
- bay area (1)
- beats (1)
- book search (1)
- bunnie huang (1)
- car dealers (1)
- carlos slim (1)
- catastrophe bonds (1)
- cdos (1)
- ces 2008 (1)
- chance (1)
- children (1)
- cochran-harpending (1)
- computers (1)
- cpi (1)

- [david x. li](#) (1)
- [dick cavett](#) (1)
- [dolomites](#) (1)
- [drugs](#) (1)
- [dune](#) (1)
- [eharmony](#) (1)
- [epidemics](#) (1)
- [escorts](#) (1)
- [faces](#) (1)
- [fads](#) (1)
- [favorite posts](#) (1)
- [fiber optic cable](#) (1)
- [francis crick](#) (1)
- [gary brecher](#) (1)
- [gizmos](#) (1)
- [greece](#) (1)
- [greenspan](#) (1)
- [hypocrisy](#) (1)
- [igon value](#) (1)
- [iit](#) (1)
- [inflation](#) (1)
- [information asymmetry](#) (1)
- [iphone](#) (1)
- [jack kerouac](#) (1)
- [jaynes](#) (1)
- [jfk](#) (1)
- [john dolan](#) (1)
- [john kerry](#) (1)
- [john paulson](#) (1)
- [john searle](#) (1)
- [john tierney](#) (1)
- [jonathan littell](#) (1)
- [las vegas](#) (1)
- [lawyers](#) (1)
- [lehman auction](#) (1)
- [les bienveillantes](#) (1)
- [lowell wood](#) (1)
- [lse](#) (1)
- [mcgeorge bundy](#) (1)
- [mexico](#) (1)
- [michael jackson](#) (1)
- [mickey rourke](#) (1)
- [migration](#) (1)
- [money:tech](#) (1)
- [myron scholes](#) (1)
- [netwon institute](#) (1)
- [networks](#) (1)
- [newton institute](#) (1)
- [nfl](#) (1)
- [oliver stone](#) (1)
- [phil gramm](#) (1)
- [philanthropy](#) (1)
- [philip greenspun](#) (1)
- [portfolio theory](#) (1)
- [power laws](#) (1)
- [psychology](#) (1)
- [randomness](#) (1)
- [recession](#) (1)
- [sales](#) (1)
- [singapore](#) (1)
- [skype](#) (1)
- [standard deviation](#) (1)
- [star wars](#) (1)

- [starship troopers](#) (1)
- [students today](#) (1)
- [teleportation](#) (1)
- [tierney lab blog](#) (1)
- [tomonaga](#) (1)
- [tyler cowen](#) (1)
- [venice](#) (1)
- [violence](#) (1)
- [virtual meetings](#) (1)
- [war nerd](#) (1)
- [wealth effect](#) (1)

G Analytics

statcounter

Simple theme. Powered by [Blogger](#).