Disclaimer: The following article is a work of parody. The presence of self-selection bias and massive sample variance (not to mention the fact that the author's credentials as a statistician amount do "doing pretty good in sixth form Maths") imply that all conclusions drawn from the results below are invariably false. In fact they are intentionally absurd.

A rigorous quantitative analysis of the jelly bean jar estimation ability of RSS stall visitors

Joshua Krissansen-Totton

Reviewers: Kevin Keirle, Natasha Turnbull, Richard Wilde

Abstract

During first week of semester, the Reason and Science Society held a jelly bean estimation ~~experiment~~ competition. Participants were asked to estimate the number of jellybeans contained in a 1.5L jar, using whatever computational method they felt was appropriate. Preliminary results suggest that RSS stall visitors were adept at jelly bean estimation. The mean estimate, 674±280, was well within one standard deviation of the true number of jelly beans, 707. Intriguingly, there seems to be evidence of small differences in estimation accuracy between different groups. Possibly sociological and evolutionary explanations for these differences are explored.

Introduction

Following the recommendations of Wilde (1996) and Keirle (2003), the Reason and Science Society held a jelly bean estimation competition during the first week of semester. Each participant was asked to estimate the number of jelly beans in a 1.5L jar to the best of their ability, using whatever method they deemed appropriate (except opening the jar and counting the jelly beans). Participants were charged $0.50 for each guess with the assurance that the person with the closest estimate would win the jar.

The different methods of jelly bean computation used by participants can be summarized as follows:

1) The cylindrical cross-section approach (total number of jelly beans given by J_T):

$$J_T = \frac{n_b h}{\Delta b}$$

Where n_b is the number of jelly beans visible at the base, Δb is the width of this bottom layer and h is the height of the jar.

2) The jelly bean density approach:

$$J_T = \frac{V(1-n)}{v_j}$$

Where V is the volume of the jar, v_j is the estimated volume of a single jellybean and n is the fraction of the jar's volume that is believed to be empty space.

3) The "I have no idea so I'll just copy the person above me" heuristic:

$$J_T = J_{T-1} + 2$$

4) The "stare at it for twelve minutes then take a wild guess" heuristic:

$$J_T(t) = \begin{Bmatrix} \dots & ;t < 720 \\ 500 \times Random(0,1) & ;t \geq 720 \end{Bmatrix}$$

5) The imaginary jelly bean conjecture[1]:

$$J_T = 4.5 + i$$

Each participant was asked to record their name, email address and jelly bean estimate. Unfortunately we lacked the foresight to record each individual's gender at the time of their estimate. Thus we were forced to guess each person's gender on the basis of their maiden name. We apologize to anyone who we may have misclassified. Half a dozen or so ambiguous cases were omitted from the analysis.

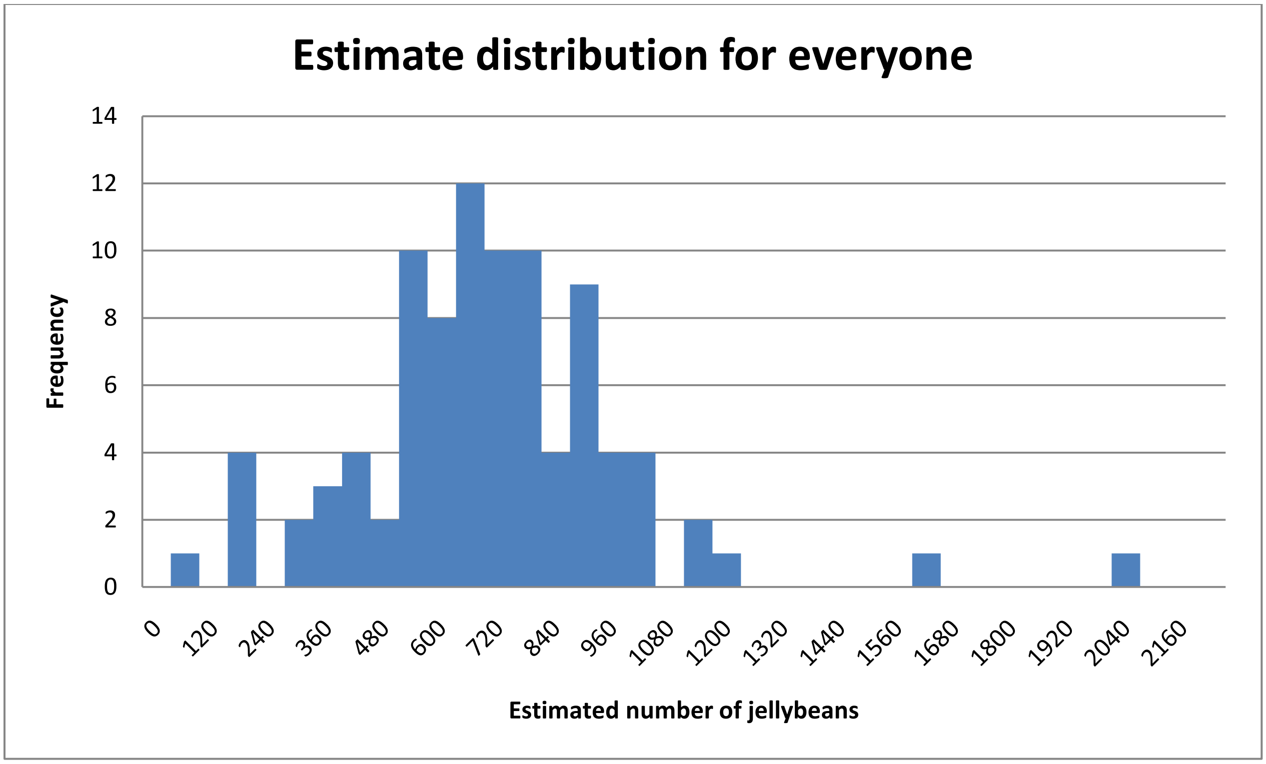
Results:

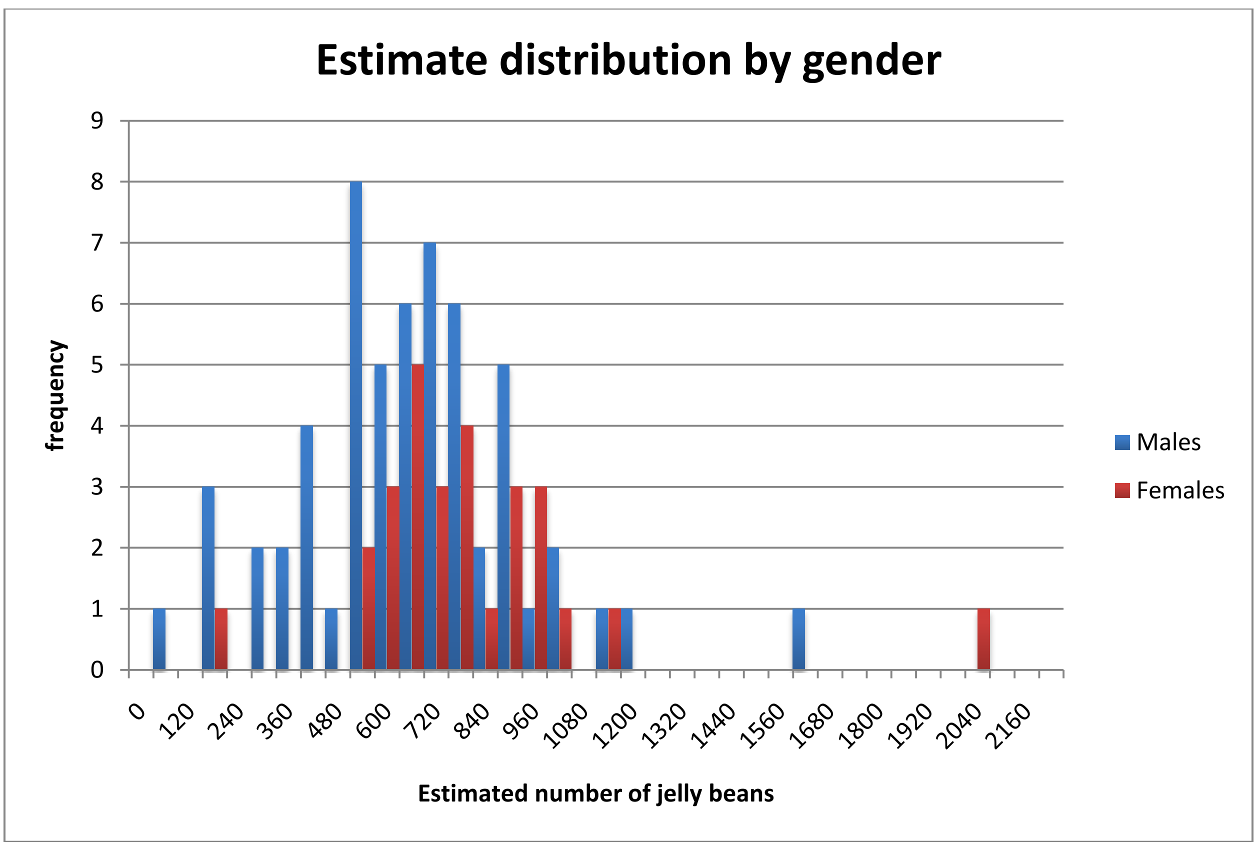
Congratulations to our winner who correctly guessed that the jar contained 707 jelly beans. The mean jelly bean estimate was 674±280 - well within one standard deviation of the true number. Descriptive statistics for jellybean estimation are shown in the table below:

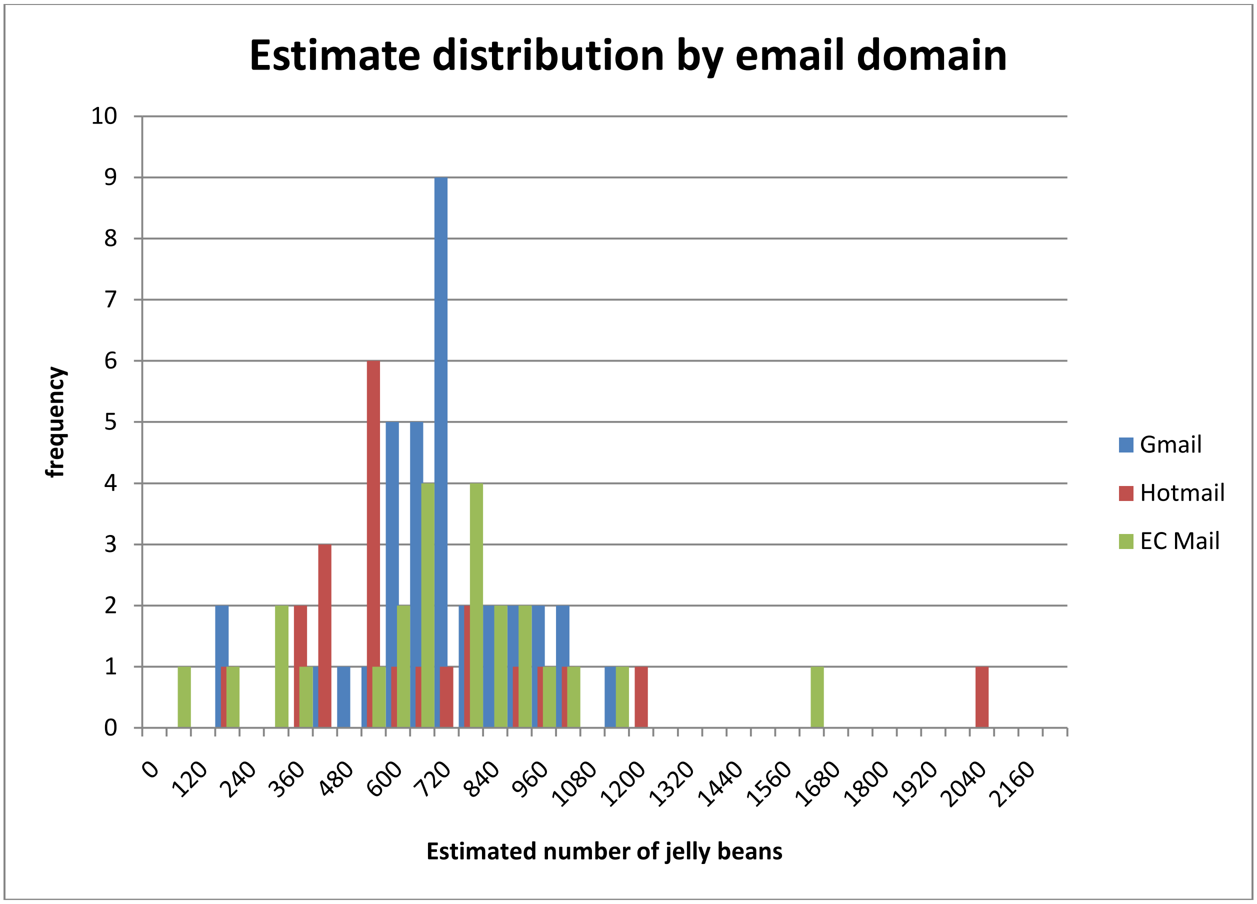
	Everybody	Females	Males	Gmail users	Not Gmail users	Hotmail users	EC mail users
Mean Jelly Bean Estimate	674.43	767.29	629.06	679.97	671.03	652.95	672.90
Standard Error	29.62	57.77	35.38	34.28	43.16	82.41	67.28
Median	663.50	729.50	646.50	687	645	522.5	692.5
Mode	500.00	#N/A	500.00	600	500	500	#N/A
Standard Deviation	284.11	305.68	269.44	202.83	325.83	386.52	329.59
Sample Variance	80721.10	93442.36	72595.26	41139.73	106163.11	149400.43	108631.48
Kurtosis	4.98	9.40	1.92	1.62	4.39	6.18	1.53
Skewness	1.11	2.20	0.49	-0.61	1.32	2.13	0.33
Range	1984.50	1839.00	1563.50	957	1984.5	1839	1563.5
Minimum	4.50	150	4.5	140	4.5	150	4.5
Maximum	1989.00	1989	1568	1097	1989	1989	1568
Sum	62047.50	21484	36485.5	23799	38248.5	14365	16149.5
Sample Size	92.00	28.00	58.00	35	57	22	24
Confidence Level (95.0%)	58.84	118.53	70.84	69.67	86.45349	171.37	139.17

The distribution of guesses is shown in the figure below. Figures are also shown comparing the guess distribution across genders and different email domains.

[1] Yes this was an actual guess. We thought the best thing to do was take the real component, although in retrospect perhaps we should have taken the modulus.


Estimate distribution for everyone
Frequency
14
12
10
8
6
4
2
0
0
120
240
360
480
600
720
840
960
1080
1200
1320
1440
1560
1680
1800
1920
2040
2160
Estimated number of jellybeans


Estimate distribution by gender
frequency
9
8
7
6
5
4
3
2
1
0
0
120
240
360
480
600
720
840
960
1080
1200
1320
1440
1560
1680
1800
1920
2040
2160
Estimated number of jelly beans
Males
Females



The mean jelly bean estimates for males and females are approximately equidistant from the true value, thus implying that neither gender is markedly better at estimating jelly bean numbers. With that said, the groups do differ in the sense that females seems to systematically overestimate jelly bean numbers whereas males seem to systematically underestimate jelly bean numbers.

The observant reader will have noticed that, contrary to what we have just stated, there is no difference between the male mean estimate and the female mean estimate according to the standard metric of statistical significance. However the Non-Gaussian features of estimate distributions imply that standard measures of significance are inappropriate in this instance. To test for significance we need to construct the estimate distributions using a Discrete Fourier Transform. If *f(x)* denotes the sample of female estimates and *m(x)* denotes the sample of male estimates (both discrete sets) then the associated distributions are given by:

$$F(x) = \int_{-\infty}^{\infty} f(x) e^{-i2\pi f t} \, dt$$

$$M(x) = \int_{-\infty}^{\infty} m(x) e^{-i2\pi f t} \, dt$$

where *f* is the frequency of sampling. Given these continuous distributions we are in a position to evaluate the 95% confidence interval for each distribution by solving the following equation for the limits of integration[2]:

$$\int_a^b F(x)\,dx = 0.95$$

and similarly for M(x) with its limits of integration c and d. Note that in order to solve for a, b, c and d we need make use of the following expression (where x-bar is the mean value of the distribution):

$$skewness = \frac{2\sqrt{(21\bar{x}\pi(a-b))}^{\,2}}{\pi\bar{x}\left(\sqrt{a}-\sqrt{b}\right)\left(\sqrt{a}+\sqrt{b}\right)}$$

Finally, given the 95% confident intervals for both the male and female distribution we can test to see if there is a statistically significant difference between the two distributions by applying the following result:

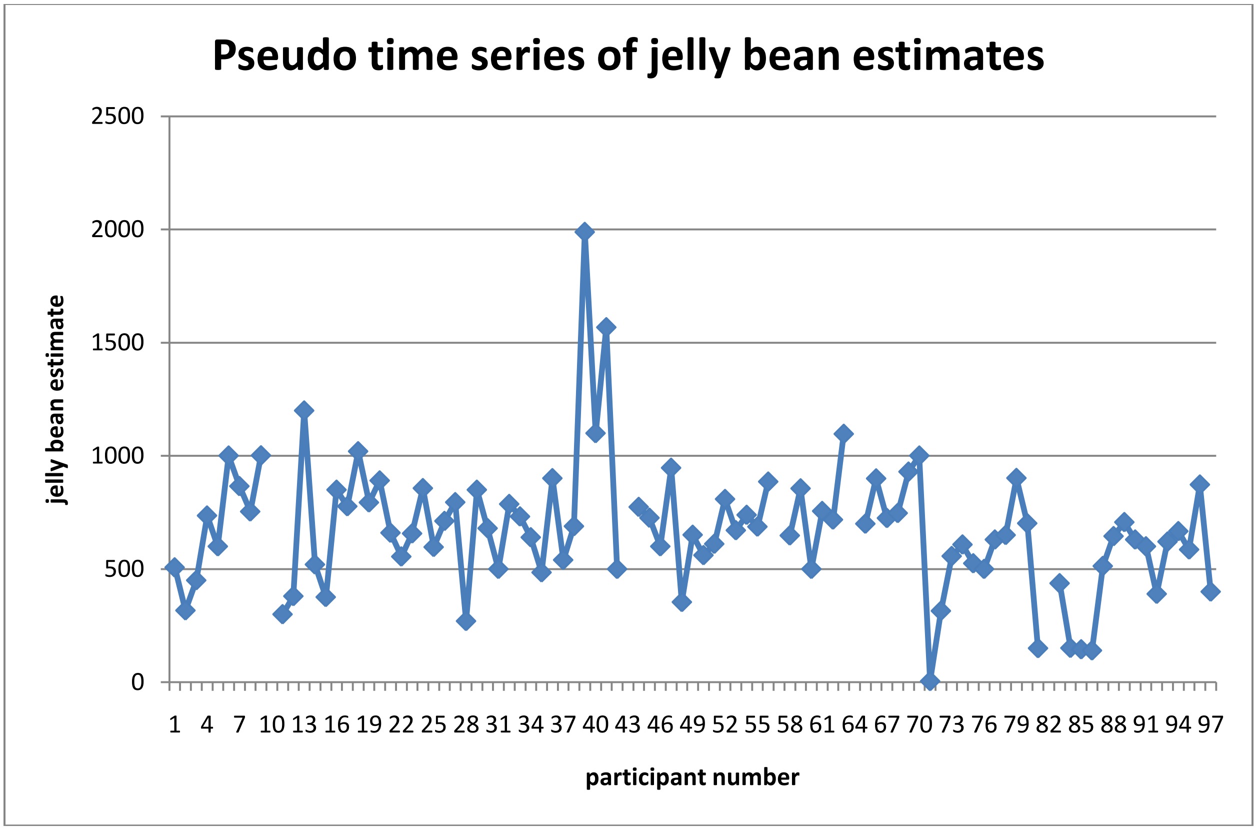
$$Estimate\ distributions\ differ\ significantly\ from\ one\ another\ if\ and\ only\ if$$

$$\sum_{n=1}^{\infty} ab\pi\, e^{cn} e^{-dn}\beta/\alpha > 9\times 10^3$$

Where α is the fine structure constant and β is the number of days until the apocalypse (May 21st 2011 - see Camping (2011) for a biblical derivation). By using this measure we deduce that the difference between the male mean estimate and female mean estimate is statistically significant. Using a similar approach it can be shown that the differences between different email domains are also significant.

As an interesting aside we plotted the jelly bean estimates in the order that they were written on the sheet to see if there was any evidence for method 3 being used by participants. Readers can make up their own minds on this issue.

[2] Yes, that's right, my normalization is so powerful that it's not necessary to talk about it.



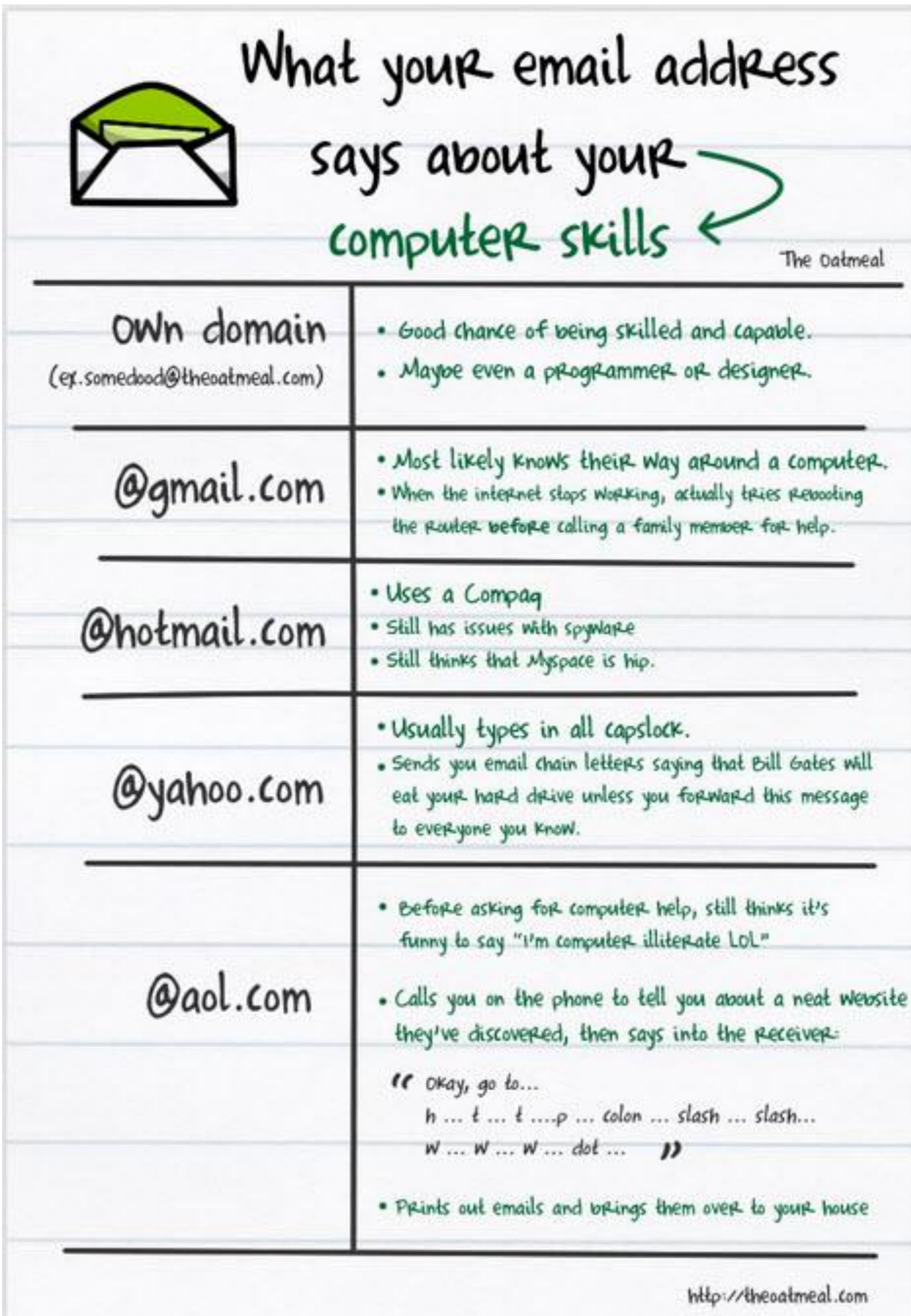
Discussion

The observation that males tend to underestimate jelly bean numbers whereas females tend to do the opposite demands an explanation. Turnbull (2010) has suggested that the tendency among females to overestimate jelly bean numbers stems from sociological pressures relating to sexual conduct. In particular, Turnbull conjectures that males often attempt to persuade female partners that the dimensions of their own genitals are vastly greater than their actual size. Consequently, in their attempts to be charitable, many females adopt a habit of overestimating length. It is conceivable that this habit of length overestimation carries over to non-sexual situations, thereby explaining the slight tendency of females to overestimate the correct number of jelly beans.

Although it succeeds in explaining the female estimate distribution, this explanation does not account for the tendency of males to underestimate jelly bean numbers. Intuitively we would expect that males accustomed to overestimating their own endowment might overestimate the number of jelly beans in the jar. However results show the opposite to be true. Wilde (2010) has explained this counter-intuitive result in terms of an ego preservation mechanism. He argues that many males may have developed a tendency to underestimate the dimensions of objects in their external environment whilst not applying the same standards of measurement to themselves. This would explain the slight tendency of males to underestimate jelly bean numbers.

Another possible explanation for the gender asymmetry emerges from evolutionary psychology (Turnbull, 2006). According to this picture, the differing estimation biases in males and females are a reflection of the division of labour in ancestral hunter-gatherer societies. If female foragers had a slight tendency to underestimate berry numbers, then the berry surplus that resulted would sustain the tribe when other food sources became scarce. This would benefit the tribe's, and by extension the female's, chances of survival. Conversely if male hunters had a slight tendency to overestimate prey numbers, this might compel them to take more risks in the pursuit of food. This would ensure a more reliable food supply for the tribe, which would ultimately benefit the male's survival chances. Of course this view implies that modern females mistake jelly beans for berries whereas modern males mistake jelly beans for TEN TONNE MASTODONS, but this is relatively a minor point of contention. More crucially this evolutionary view has been rejected by Dawkins (1993) because it hints of group selection, and is therefore a load of creationist nonsense. Furthermore, this position has also been strongly criticized by Krissansen-Totton (12 seconds ago) on the basis that it's the wrong way around.

In explaining the difference in jelly bean estimation ability between different email users, we appeal refer to the work of Oatmeal et al. (2006) which is summarized below.



Conclusions
We believe we can draw the following tentative conclusions from the results presented above:

- People are generally fairly good at estimating jelly bean numbers.
- Males think things are smaller than they actually are whilst females think things are bigger than they actually are. Go figure.
- Gmail users are naturally superior to everyone else (except those people who create their own domain names).
- RSS should have a Mastodon counting competition.
- The RSS secretary clearly has too much time on his hands.

Acknowledgements
Thanks again to everyone who helped out at the RSS stall – your help was very much appreciated. Also thanks to Betty for being awesome.

References (in order of publication)

Wilde, R., 1996. On the desirability of jelly bean jars. Oxford University Press, Oxford.

Keirle, K. 2003. Every jelly bean is the tear of a unicorn: what jelly beans can tell us about society, and our childhood. Auckland University of Technology University press, Auckland.

Wilde, R., Keirle, K. 2005. An exhaustive search algorithm for determining the true number of jelly beans in a jar. Advances in Mathematics 120, 340-357.

Turnbull, N. 2010. Why males mistake jelly beans for Mastodons. Journal of Anthropology, 23, 130-169.

Camping, H. 2010. Repent or burn, bitches: a rigorous derivation of the date and time of the apocalypse. Journal of Biblical Theology 58, 34-78.