JWT's copy
(11 extra copies in file - 3/5/70)

SECOND DRAFT--NOT FOR RELEASE
UNDER ANY CIRCUMSTANCES

STATEMENT ON THE DESIRABLE FORM
OF A DRAFT-LOTTERY DRAWING

Opinions as to the overall fairness of any draft system always seem to vary widely. I am here to discuss only one very narrow aspect--how to conduct a lottery--so that this one aspect will be as fair as possible.

A drawing for a draft lottery should be designed to meet three main principles:

- the greatest practical fairness to each draft registrant.

- responsiveness to the well-known British legal maxim that justice should not only be done but be seen to be done.

- recognition that some appeal to the principle of de minimis non curat lex is essential, but that possible imperfections in the drawing process should be kept at a very low level indeed.

To meet these principles adequately, it would be wise to carry out a more careful drawing than has so far been used--so far as I am aware--either for the selection of draftees or for the selection of sweepstakes or lottery winners or for the selection of bonds for compulsory redemption. Our knowledge of random processes, both in theory and in practice, is now such that it would not be difficult to meet such a new high standard.

Before describing one or two overall processes that could be used to meet such a standard, I should say a few simple things about randomization in general.

Randomness is in the process, not in the result.

From the point of view of the statistician and of the experienced user of random processes for scientific purposes, much of the criticism of the recent draft lottery has not been thought through sufficiently carefully. Randomness does not lie in the appearance of a single result. Rather, it lies in our understanding and knowledge about the relative probabilities of the possible outcomes --this comes only from knowledge and understanding of the process.

Suppose it is agreed that some decision is to be decided "on the toss of a coin". Suppose we know that a coin was "tossed" and it came "heads". Can we judge from this single result--without knowledge of the coin and the tossing process--whether or not this toss was random? Surely not.

The "toss" was random for a fair coin if "head" and "tail" were equally likely to arise. A single "toss" offers no evidence as to whether or not this was true. We can, for example, from the single observed "head" rule out the possibility that the coin used had two tails, but we cannot rule out the possibility that it had two heads.

In arranging for our coin "toss" to be random, we may have to be careful of rather subtle physical or psychological effects. If we are told that the "toss" was conducted by spinning an unworn silver dollar on a very smooth sheet of glass, the spin seems very fair until we learn that the shape of a silver dollar is such that the probability of "heads" under such circumstances is very much greater than one-half.

In striving for randomness, and in assessing the quality of randomness that has been achieved, we need to give our attention to the process. The only importance of individual results lies in the light they throw on the process, not in their own nature as such.

In the case of a drawing to do something as complex as arranging dates in a supposedly random order, we often wonder whether the results support the hypothesis that the drawing process was statistically random or the hypothesis that mixing is incomplete. In extreme cases, the results of a single complex drawing could give serious evidence favoring one of these hypotheses in comparison with the other.

Fairness is enhanced by combination of independent nearly-fair processes.

If two parties wish to have a nearly fair "toss of a coin" what can they do? The simplest thing is for each to toss a coin and let the result rest on whether the coins match of not. Here the first tosser will ensure independence as a matter of policy. If either tosses his coin with exactly 50-50 chance of heads or tails, the probability of a match is exactly 1/2 no matter what the chance of heads is in the other tossing. The probability of a match is still 1/2, for instance, even if the other party is tossing a two-headed penny. The details are illustrated on Exhibit 1.

More is true, however, and has great practical importance. If both tossers come within even 1% of 50-50 probability--that is between 49% and 51%--and if there is no connection at all between their tossing processes, so that their tosses are indeed "independent" in the sense that that term is used in the science and mathematics of probability, then the probability of a match must come within two-hundredths of 1% of 50-50--that is between 49.98% and 50.02%. An extreme case is illustrated in Exhibit 1.

This great increase of fairness from the combination of independent events is well-known and well-understood. It has, for example, been used to improve the quality of large printed tables of random digits, and is routinely used in scientific applications of randomness.

There is no need to confine the application to only two independent processes. If a neutral assistant is brought in, he may, for example, toss a third coin with the agreement that a "match" between A and B wins for A if the third coin falls "heads" but for B if the third coin falls "tails", while a failure to match has the opposite consequences. Exhibit 2 shows some details. If all three coins come within 1% of being 50-50--between 49% and 51%--and are independent, the chances of either party winning must now be within four ten-thousandths of 1% of 50-50.

Such high precision is rarely necessary. Situations where it is worthwhile to combine three independent processes can arise, however, when the individual processes are quite crude. If each toss is only known, for example, to have a probability of "heads" between 40% and 60%, three independent tosses combine to give a probability between 49.6% and 50.4%. Thus even crude probabilities can be rapidly brought into line by combination--if we can ensure independence.

Tables of random numbers.

Statisticians are often concerned with the results of clearly defined random processes. They use the behavior of such ideal processes as guides in dealing with the more-or-less non-ideal processes of the real world. Some aspects of these ideal processes can be most efficiently studied by the manipulation of mathematical formulas. Others would involve too much mathematical effort, and are better studied by simulating the ideal situation experimentally. Like all other physical simulations, such experimental simulations will not be quite perfect. It has been worthwhile to statisticians to spend considerable thought and effort on improving the quality of randomness in such simulations.

By 1925 L. H. C. Tippett, later an eminent British statistician and director of important textile research organizations, despaired of meeting sufficiently high standards of quality by mixing objects in a bowl and drawing from it. He found, incidentally, that he could do as well by mixing metal-rimmed price tags as with any other objects he tried, but he found this not good enough for his needs.

As a result he started with apparently randomly arranged numbers, subjected them to various processes of shuffling and re-arrangement, and published the result as the first table of random numbers.

During the intervening four decades, a substantial number of such tables have been prepared, and tested both by use and by searching investigation. They are now regarded by the profession as one of its standard tools, and have essentially completely replaced the use of physical shufflings or stirrings as a basis for either theoretical work or the practical conduct of random sampling.

<u>Precautions to be used in random drawings from a bowl.</u>

Those not professionally trained in statistics need not have --and should not be expected to necessarily have--the confidence in random number tables that statisticians have learned from experience to have. Thus there remains a place for drawings from a bowl as at least part of a public random lottery. While statisticians have been unsuccessful in making drawings from a bowl sufficiently random for their most delicate and refined purposes, they have learned quite a bit about useful precautions in using bowl-drawing.

It would not be appropriate for me to try to list today all the precautions that might reasonably be taken in order to make bowl-drawing quite close to randomness, but to give a few will be useful by indicating a general flavor and approach.

1) The advantages of making the objects to be stirred as similar in size, weight, smoothness, and appearance as possible are widely recognized.

2) The advantages of not only stirring the objects before the first draw, but also further rearranging them--perhaps by turning the vessel in which they are contained, perhaps by stirring--between each pair of draws have long been recognized.

3) There are significant advantages to placing the objects into the "bowl" at the beginning in a pre-chosen well-shuffled order rather than in a systematic order. This can help greatly in overcoming difficulties with stirring. (While some statisticians apparently feel that this would be enough to quiet all fears, I should like to see even more precaution, as explained below.)

4) Those conducting the draw need experience in the effectiveness --or lack thereof--in their procedures of stirring, tumbling and the like. They will do well to conduct pre-experiments, starting, for example, with one-half of the bowl filled with black objects and the other half filled with white ones, and learning by experience how much stirring is required to produce a reasonably uniform appearance.

Arranging that "justice be seen to be done".

The advantages of combination of independent nearly-random processes can easily be had in the type of drawing used in a draft lottery. Our overall purpose in the main draw is to set up a matching between the 366 dates--1 January through 29 February to 31 December--and the numbers from 1 to 366. We can easily do this in two stages, making one matching between the dates and the numbers from 501 to 866, and a second matching between the numbers from 501 to 866 and the numbers from 1 to 366.

Let me illustrate what I mean on a simpler example, matching the names of days of the week with numbers from 1 to 7. We will do this in two stages, matching the names of the days to the numbers from 51 to 57 and separately matching the numbers from 51 to 57 with the numbers from 1 to 7. Exhibit 3 shows the procedure.

Returning to the real problem, we can combine our two matchings in an exactly similar way. Thus if 16 June goes with 864 in the first matching while 864 goes with 137 in the second, the combination assigns 16 June to 137. Such combination is simple, and can be understood by the general public.

The combination of two very careful bowl-drawings, perhaps carried out simultaneously in widely-separated cities, would be much more random and much fairer than the results of any one bowl-drawing. Such a choice--even if conducted with the care hinted at by the four precautions mentioned above--would, however, not be immune from professional criticism.

In my personal judgment, very careful consideration should be given to the combination of two matchings of very different character, one involving the use of tables of random numbers, the other physical drawing from a bowl.

One might, for example, match the calendar dates to numbers from 501 to 866 by a bowl drawing, while independently--perhaps at a distant point--using tables of random numbers to match the numbers from 501 to 866 with the numbers from 1 to 366. In the latter matching use might also be made of the closely related tables of random permutations.

By combining matchings in this way, both the general public and professional critics would be able to see that justice was indeed being done.

<u>Remark on the use of tables of random numbers.</u>

There are various ways in which a table of random numbers, starting at a chosen point, can be used to provide the sort of matching required--for example, a matching of the numbers from 501 to 866 with those from 1 to 366. It will be important to have the chosen process carefully examined, in advance, by competent statisticians and mathematicians. Subtle biases can arise far too easily; there is no substitute for expert judgment--in advance.

The problem of choosing a start is simple enough, however, to be worth a few words here. Suppose the table in question involves a million starting points--some do. Our problem is to select, as randomly as maybe, a number between 000000 and 999999.

be
can (?)

Different people will have different processes they favor: shuffling packs of cards, rolling 10-sided dice, opening other tables of random numbers "at random", etc. These can all be accommodated, and we can have--at the same time--the gains in randomness from combining independent processes. We have only to allow as many people as we wish to prepare, and place in a sealed envelope a number between 000000 and 999999. The envelopes can then be opened, the numbers added together, the millions discarded, and the right-hand 6 digits of the sum used to fix a starting point.

If any one person were to pick his number exactly at random, independently of the others, the resulting starting point will be perfectly at random. If any two people nearly meet this standard, the result will be random for all practical purposes.

Close.

My purpose here today has been to explain to you four things:

- that randomness is in the process, not in a specific realization.

- that combination of independent processes is often the key to highly precise approximation to perfect randomness.

- that statisticians have had to give up the stirring of bowls as the source of the best randomness, but that such bowls--when very carefully used--may play a significant role in convincing the general public that an overall process is indeed random.

- that we know enough about nearly-random processes to attain a very high degree of randomness indeed, far better than has been reached in any drawing--but only if we exercise extreme care throughout.

So far as the draft lottery alone is concerned:

- we can be extremely fair to every draft registrant.

- we can arrange for justice not only to be done, but be seen to be done.

- we can make our deviations from ideal randomness smaller than ever before.

- we have the knowledge; we have only to apply it with the extreme care our problem deserves.

2 March 1970

John W. Tukey

Exhibit 1

Coin tossing by matching

A) The POSSIBILITIES -- numerical probabilities in panels B to E

A's toss:	HEADS	HEADS	TAILS	TAILS
B's toss:	HEADS	TAILS	HEADS	TAILS
Winner:	A	B	B	A

B) TWO FAIR COINS -- probabilities for INDEPENDENT tosses

A's toss:	50%	50%	50%	50%
B's toss:	50%	50%	50%	50%
Combined:	25%	25%	25%	25%

	A	B
Total chances of winning:	25%	25%
	25%	25%
	50%	50%

C) TWO QUITE UNFAIR COINS -- probabilities for INDEPENDENT tosses

A's toss:	40%	40%	60%	60%
B's toss:	55%	45%	55%	45
Combined:	22%	18%	33%	27%

	A	B
Total chances of winning:	22%	18%
	27%	33%
	49%	51%

D) ONE FAIR and ONE UNFAIR COIN -- probabilities for INDEPENDENT tosses

A's toss:	50%	50%	50%	50%
B's toss:	10%	90%	10%	90%
Combined:	5%	45%	5%	45%

	A	B
Total chances of winning:	5%	45%
	45%	5%
	50%	50%

E) TWO NEARLY FAIR COINS -- probabilities of the 4 patterns

A's toss:	49%	49%	51%	51%
B's toss:	49%	51%	51%	49%
Combined:	24.01%	24.99%	26.01%	24.99%

	A	B
Total chances of winning:	24.01%	24.99%
	26.01%	24.99%
	50.02%	49.98%

Exhibit 2

Coin tossing by matching with an assistant

A) The POSSIBILITIES -- numerical probabilities in B and C

A's toss:	HEADS	HEADS	HEADS	HEADS	TAILS	TAILS	TAILS	TAILS
B's toss:	HEADS	HEADS	TAILS	TAILS	HEADS	HEADS	TAILS	TAILS
Assistant:	HEADS	TAILS	HEADS	TAILS	HEADS	TAILS	HEADS	TAILS
Winner:	A	B	B	A	B	A	A	B

B) THREE quite UNFAIR COINS -- probabilities for INDEPENDENT tosses

A's toss:	60%	60%	60%	60%	40%	40%	40%	40%
B's toss:	60%	60%	40%	40%	60%	60%	40%	40%
Assistant:	30%	70%	30%	70%	30%	70%	30%	70%
Combined:	10.8%	25.2%	7.2%	16.8%	7.2%	16.8%	4.8%	11.2%

Total chances of winning:

A	B
10.8%	15.2%
16.8%	7.2%
16.8%	7.2%
4.8%	11.2%
49.2%	50.8%

C) THREE NEARLY FAIR COINS -- probabilities for INDEPENDENT tosses

A's toss:	51%	51%	51%	51%	49%	49%	49%	49%
B's toss:	51%	51%	49%	49%	51%	51%	49%	49%
Assistant:	51%	49%	51%	49%	51%	49%	51%	49%
Combined:	13.2651%	12.7449%	12.7449%	12.2451%	12.7449%	12.2451%	12.2459%	11.7649%

Total chances of winning:

A	B
13.2651%	12.7449%
12.2451%	12.7449%
12.2451%	12.7449%
12.2451%	11.7649%
50.0004%	49.9996%

Exhibit 3

Combining two "random" matchings

A) The FIRST MATCHING--one "random" draw

Thursday	51
Monday	52
Sunday	53
Friday	54
Tuesday	55
Wednesday	56
Saturday	57

B) The SECOND MATCHING--another "random" draw

53	1
52	2
57	3
51	4
54	5
56	6
55	7

C) The COMBINED MATCHING

Monday	52	2
Tuesday	55	7
Wednesday	56	6
Thursday	51	4
Friday	54	5
Saturday	57	3
Sunday	53	1