ELICITATION RECORD – Part 2 – Distribution

Roulette Method

Elicitation title	As in the Part 1 form
Session	As in the Part 1 form
Date	As in the Part 1 form
Quantity	The uncertain quantity whose distribution is to be elicited
Start time	Time when this part of the elicitation started

Definition	Repeat the definition of this quantity from Part 1. Give it a symbol to facilitate the recording of judgements about it. It will be called X in these notes.
Evidence	Review the evidence specifically about X. (Refer to principal sources, but do not repeat lots of detail here.) [Although the evidence base has been set out in the Part 1 form, the facilitator should ask each expert to consider which items are of relevance to this quantity.] *As in Part 1, this step is to avoid the 'availability heuristic', in which experts rely only on a subset of evidence that comes readily to mind.*
Plausible range	Elicit from the experts the range of plausible values for X. In these notes, L is the lower bound of the range and U the upper bound. It may be useful to record absolute, logical bounds, but the objective here is to identify a range such that it is extremely unlikely (but not necessarily impossible) that X lies outside. Record how the range has been informed by the evidence base. [This range should not be unnecessarily wide, but it is important that it should not be too narrow. This should be a joint judgement of the experts, such that they all believe that X is extremely unlikely to be outside (U, L). After an initial specification of the range, the facilitator should probe by asking something like "Suppose an experiment produced a value [something below L or above U] for X; would this have to have been a flawed experiment, or might there be a way that X could have this value?"] *There is substantial evidence that experts tend to be over-confident, in the sense that they do not allow enough probability for extreme values of X. Psychologists suggest that experts develop mental models that allow them to understand and predict the processes in their area of expertise. Experts themselves may recognise instances where the model does not*

	apply, but part of their overconfidence in elicitation may be due to conditioning on inappropriate models. The suggested probing question invites them to think outside their models.
Chips in bins	The next step should be done by each expert **separately**, without discussion. The facilitator divides the range from L to U into 10 equal-width 'bins'. Each expert should specify their probabilities for X to lie in each of the 10 bins by placing 'chips'. For example if L = 10 and U = 100, the first bin is from 10 to 19, the second from 19 to 28 and so on. The number of chips that an expert places in the second bin represents the expert's judgement of P(19 < X < 28). [The facilitator may find it helpful to adjust the L and/or U values so that the bin boundaries have more rounded values. Experts should have pre-printed sheets with ten bins marked out, and the facilitator will instruct them to write in the boundaries. Experts will also have been given a number of chips each to place. This number is at the facilitator's discretion. A typical number might be 20, in which case the facilitator may advise the experts that each chip represents a probability of 0.05. Alternatively, 25 chips would each represent 0.04, while 10 chips would each represent a 0.1 probability. The facilitator should not use fewer than 10 chips; more chips in principle leads to more accurate probability specifications by the experts, but takes more time and in practice a sensible upper limit is 30. The facilitator may advise the experts that a realistic expression of uncertainty should involve concentrating chips in relatively few bins, but not too few. They should normally use at least 3 different bins. Also, unless L and/or U are physical bounds, the fact that it is considered implausible for X to be outside these bounds suggests that the extreme bins (with L or U as one of their boundaries) would not have sufficient probability to receive even one chip. When all the experts have placed their chips, the facilitator should ask them to say how many they have placed in each bin.] *In the Roulette method the experts can adjust their deployment of chips until they are satisfied with the distribution. We recommend using actual physical chips (rather than just writing a number of chips in each bin or drawing blobs), because this engages the experts well and allows a visual as well as a numerical representation of their uncertainty.* *Psychology research identifies the range-frequency compromise as a tendency for experts to spread probability evenly over the available options, which would lead to unrealistically high uncertainty when using the roulette method. This is the reason for the facilitator advising the experts about concentrating the chips in a few bins.*

Fitting	The facilitator fits a distribution to each of the experts' assessments. (The distributions should be specified here, and if possible shown as density functions. If it is not straightforward to paste plots of the density functions into this record, they can be provided as an attachment, which should then be listed at the end.) [The facilitator should choose an appropriate family of distributions, and then fit the distribution by choosing parameters that give probabilities matching the elicited bin probabilities. The distributions should be shown to the experts, but at this stage we do not invite revision (unless the expert is insistent that the plotted distribution badly distorts his/her beliefs) or provide any other feedback. If feasible, the facilitator should compute the median and tertiles of an equally-weighted average of the density functions. Ideally, these should not be revealed to the experts immediately, but may be used at the facilitator's discretion in the next stage.] *This stage of separate elicitations ensures that the initial divergence of opinion between the experts is recorded. The facilitator can refer to these if the group elicitation appears to be neglecting part of the original range of belief. There is evidence that group elicitation can itself lead to over-confidence, perhaps because the process of reaching consensus induces a false sense of decisiveness. So this step in the SHELF process allows the facilitator to see any narrowing of uncertainty, and to check that this is justified by the sharing of knowledge that has taken place. In particular, having the median and tertiles of the averaged densities gives the facilitator a quick check on the degree to which subsequent discussion leads to a narrowing of uncertainty.* *The process of averaging the density functions is known as the linear opinion pool (with equal weights). It is one of the formulae which proponents of eliciting separately from experts use to combine the resulting distributions. We use it in SHELF simply as a benchmark.*
Group elicitation	After discussion of the different distributions, and sharing of knowledge and reasoning about the differences, create and record a group allocation of chips to bins. The same set of 10 bins should be used, with the facilitator placing chips on a single sheet. The placement should represent the experts' group judgement about probabilities. [The way in which the chips are placed in this stage is at the discretion of the facilitator. One approach is to begin with the chips evenly allocated over the bins and to invite the experts to

	move chips from the edges (less probable bins) towards more probable bins. Another is to start with a blank sheet, ask for agreement on the most probable bin and allocate a few chips there, then to ask for a next most probable bin and so on. The discussion can take a substantial amount of time. The facilitator should only cut it short if he/she feels that the experts are no longer exchanging information and arguments but are just repeating opinions. The agreed probabilities will inevitably be some sort of compromise. Before discussion, there are two components of uncertainty in the group – uncertainty that each expert has and is expressed in that expert's quartiles, as well as variability between the experts' judgements. The agreed bin probabilities should reflect the group's overall uncertainty that remains after the discussion. The facilitator needs to manage the discussion so that divergent views are properly considered, and to ensure that strong personalities and/or groups of people with overlapping experiences do not inappropriately dominate. In general, we would expect the group debate and sharing of opinions to result in a distribution that is narrower than the linear opinion pool, but it should not be much narrower or markedly different in other ways, unless the discussion has clearly altered individual experts' opinions. In some situations, experts may not be able to reach consensus. It may then be necessary for the facilitator to elicit two (or more) distributions, representing consensus views within the opposing factions.]
Fitting and feedback	Record here the (potentially iterative) process of fitting, feedback and revision of the group judgements. [The facilitator first fits a distribution to the group's elicited probabilities. This should be shown to the experts, and the fitted probabilities compared with elicited values. The experts are invited to consider whether the fit is acceptable. The facilitator then feeds back to the experts some implied probabilities in the fitted distribution, such as the 10^{th} and 90^{th} percentiles, or the median and quartiles and tertiles. The experts are invited to consider whether these are reasonable reflections of the group's knowledge. If revision is needed, this may be followed by further rounds of fitting and feedback until the experts are comfortable with the fitted distribution and its implications.]
Chosen distribution	Record and show here the finally fitted distribution.
Discussion	The facilitator should record here any difficulties that arose during the elicitation of this distribution. Also the experts' reactions to the process and to the finally fitted distribution.

	The elicitation record should be open about any concerns with the finished distribution. The SHELF protocols are designed to avoid many of the pitfalls of elicitation, but no process is perfect. It is important to be critical and realistic about the result. Nevertheless, it is important also to remember that, despite whatever deficiencies it might have, the elicited distribution is our best attempt. It has been developed using a robust protocol, and since expert knowledge is needed in the wider enterprise there is no alternative!

End time	Time when elicitation of this distribution was completed.
Attachments	List any attachments, e.g. plots of distributions.