11 Comments on the Chapters by MacCrimmon, Stanbury, and Wehrung; and Schum

R. Duncan Luce
Department of Psychology and Social Relations, Harvard University

One risk of being a discussant at a conference is that the author takes one's criticisms seriously in making revisions so that little remains to write when preparing written remarks. To a degree that is the situation in which I find myself. My initial major comment about the chapter by Kenneth MacCrimmon, William Stanbury, and Donald Wehrung was that the most striking finding—and it certainly is that—was not sufficiently emphasized. I have in mind the fact that in two of the ranked sets of gambles, there was a pair of gambles in common, and a substantial fraction of subjects ranked them differently depending on the context. In the revised chapter, this result is given prominence, and so I am left with nothing really to say except to note that this result is enough to cast in doubt our whole current enterprise of model building in this area.

Concerning David Schum's highly interesting and informative chapter on cascaded inferences, my original comments entailed a somewhat extended discussion of a particular example, which I found very disturbing. So, I gather, did Schum, for he has examined it and a number of related examples in considerable detail, written a long memorandum about the issues involved, and prepared a paper on it, which will be published elsewhere. He alludes to these considerations in his revised manuscript for the present volume, but I have the impression that these remarks, although clear enough for those who heard my comments at the conference, will seem a bit elliptic to others. It may, therefore, not be amiss for me to repeat the example here.

I was led to consider it because, despite the fact that I had the reprints in which his equations for cascaded inference are derived, I found it difficult to sense exactly what these formidible equations said. In such a situation, it is usually wise to examine a bare-bones case that still retains the basic idea—here, that of

cascading information. Being a part-time psychophysicist, I thought immediately of a simple two-stimulus, two-response design, such as yes–no detection, in which there are independent repeated observations—say, by a set of distinct observers—that are to be aggregated into a group decision. In the usual psychophysical notation, $H_1 = s$ stands for the hypothesis that a signal (in noise) was presented and $H_2 = n$ stands for the hypothesis that no signal (noise alone) was presented. Let us identify the event D with the presentation of a signal, so in this special example:

$$P(D|H_1) = 1 \text{ and } P(D|H_2) = 0.$$

The testimony of observer i, D_i^*, is simply the observer's assertion that a signal was presented; in this context, this is called the yes response, Y. And, the testimony D_i^{c*} is the no response, N. To maintain the simplicity of the example, let us assume that all of the observers are independent and statistically identical, and so their performance is completely described by two conditional probabilities, $P(Y|s)$ and $P(Y|n)$.

From these assumptions, it is not difficult to show that Schum's Eq. (1) is (in this special case only) an uninteresting triviality and that Eq. (2) simplifies to:

$$\begin{aligned}\Lambda_F^* &= \frac{1+V}{V} \\ &= \frac{\prod_{p\epsilon P} h_p/f_p}{\prod_{q\epsilon Q} c_q/m_q} \\ &= \left(\frac{P(Y|s)}{P(Y|n)}\right)^y \left(\frac{1-P(Y|s)}{1-P(Y|n)}\right)^{n-y}\end{aligned}$$

where y is the number of observers saying Y and $n - y$ the number saying N.

What possible merit can there be to this change of notation? None—except for one thing. The psychophysical example reminds one of the very firm and important psychophysical discovery of the third quarter of this century that well-practiced, conscientious observers are not adequately characterized by a single pair of conditional probabilities, as had been implicitly and explicitly assumed during the preceding hundred years, but rather by a continuum of such pairs. The locus of such points is called the ROC curve (engineering lingo standing for receiver operating characteristic) or isosensitivity curve (psychological lingo for the same thing) or power of the test (statistical lingo). If one alters the stimulus conditions, for example, by making the signal stronger or weaker, then the ROC curve alters. But, if one holds the stimulus conditions fixed and only alters cognitive or motivational factors (for example, instructions, payoffs, presentation probabilities), then a single curve is involved and these factors determine the point actually observed. We speak of the mechanism for selecting a point on the curve as the setting of a criterion or a response bias.

I dwell on this point not because psychophysics is so intrinsically interesting, but because there is every reason to believe that this tradeoff phenomenon is very widespread whenever observers are engaged in making difficult observations for which their performance is less than perfect. That is usually the case for eye witnesses. If this is accepted, then I wish to demonstrate that the process of cascading greatly exaggerates the extent of the response bias.

Consider a symmetric, Gaussian ROC curve that has a value of $d' = 1.80$ (this is generated by noise and signal distributions that are Gaussian with unit variance and means that differ by 1.80). One pair of probabilities that lie on this curve is $P(Y|s) = .7$ and $P(Y|n) = .1$. These would arise under instructions that mildly invite the observer to be conservative when saying yes or, equally, under a payoff matrix that made the error of saying yes when there is no signal (false alarm) several times more costly than the error of failing to respond to a signal (miss). A second pair of probabilities that also lie on the same curve is .9 and .3. These would arise by instructions or payoffs favoring, to about the same degree, a more liberal criterion for saying yes. Suppose that there are 20 observers, equally split between saying Y and N. If you are dealing with observers under the first condition, then Schum's Eq. (10.2) yields:

$$\Lambda_F^* = \left(\frac{.7}{.1}\right)^{10}\left(\frac{.3}{.9}\right)^{10} = 4784;$$

whereas for observers under the second condition:

$$\Lambda_F^* = \left(\frac{.9}{.3}\right)^{10}\left(\frac{.1}{.7}\right)^{10} = 1/4784.$$

The ratio between these two likelihoods is over 22 million, a rather staggering difference when you consider that the only difference between the two cases is a relatively slight tendency to be conservative or to be liberal in saying yes.

In sum, a criterion shift that seems relatively modest to a psychophysicist—and I dare say to any student of human decision making—one that is well within the reach of judicial rhetoric, is reflected in a highly amplified form when judgments are cascaded. I am by no means sure what we should make of this fact. It is inherent in the nature of the situation, but I am sure it is important for us to be aware of it. In particular, the apparent certainty that can arise in cascaded inferences must, I believe, be closely tempered by the realization that it may reflect little more than a consistency of response bias on the part of observers. Care must be taken to try to separate the impact of response bias from that of accumulated information and to heed only the latter. Exactly how this should be done in practice is far from clear.