METHODS & DESIGN

Toward a complete analysis of GSR data¹

WILLIAM F. PROKASY, *University of Utah, Salt Lake City, Utah 84112*

A critique of amplitude analyses in GSR conditioning is discussed. It was pointed out that, when amplitudes are properly analyzed with all Ss contributing equally to each repeated measures point, the concern of latency interval, definition of response probability, and distribution of GSR amplitudes are concerns which apply to analyses of magnitude, probability, and latency, as well as of amplitude. It was concluded that nothing is lost in separating magnitude into its amplitude and probability components, particularly when magnitude, as well, is included in the statistical analyses, and that, to the contrary, we stand to learn something by such a partitioning which could not be learned without it.

Kimmel (in press) seems to object to the use of GSR amplitude and probability for a variety of reasons. The purpose of this paper is to show that the issues Kimmel raised, while related to general problems of measurement, are not related specifically to whether or not GSR amplitude and probability constitute useful measures.

Kimmel's major concern is that with repeated measures on the same S there may be S selection which results in nonrepresentative functions, the example being one in which 300 Ss receive 20 adaptation trials to a stimulus consisting of a dim white light. As his data show, the GSR decreased in magnitude, frequency, and amplitude across trials. The distortion occurred in a plotting of amplitude across pairs of trials in which an S's data were included at a data point only if S made at least one response in the two trials. If S failed to make a response on both of the trials, then no entry was made for him at that point. The result was (with Ss exhibiting a decrease in response frequency) that fewer and fewer Ss contributed to the adaptation function. As it turns out, those Ss who initially made small responses also quit responding relatively early in the adaptation series, the consequence being an "amplitude curve" which first decreased, and then increased. Kimmel points out that such a function is not "an accurate reflection of the habituation process."

What Kimmel seems not to have recognized is that the removal of data on the basis which he describes is in no way either required for, or implied by, an amplitude analysis. Moreover, in none of the papers he cites (Gormezano, 1966; Morrow & Keough, 1968; Prokasy, Fawcett & Hall, 1962; Prokasy & Ebel, 1964, 1967) is that method of recording and analysis either employed or advocated. To the contrary, Prokasy & Ebel (1967), and Morrow & Keough (1968) specifically mentioned the need to insure that the same Ss contribute to each repeated-measures data point. Since it is not uncommon in a classical conditioning situation to be concerned with measuring such analog characteristics as latency, amplitude, etc., a problem is posed when Ss do not respond on each trial. The standard solution to this problem has been to include in the analysis the data of only those Ss who meet a minimum response criterion in each block of trials. This, of course, results in the data of some Ss (i.e., those who respond infrequently) being excluded from the entire analysis; consequently, the criterion must be selected

carefully. However, the problems associated with selecting a criterion are not peculiar to GSR amplitude: they exist for any analog measurement on a response which has a likelihood of less than 1.0 in some unit of time. Therefore, S selection as a problem is not germane to the issue of whether or not it is useful to analyze GSR amplitude.

Kimmel observes that if manipulating an independent variable results in an increased latency, it is possible for a GSR to occur out of the predetermined latency range and, therefore, to be scored as a zero (or nonresponse for probability purposes). This is a general problem encountered in those situations in which Es define ahead of time what latency range is acceptable for scoring a response without knowing the range characteristic of S. The sensible solution is to define latency intervals for responses which accommodate the distributions yielded by S, rather than the a priori predilections of the E. However, if an arbitrary scoring interval is employed, and responses do occur outside of the interval, then a distortion exists for magnitude (entering a zero instead of a larger number), probability (classifying the trial as a nonresponse, rather than as a response, trial), amplitude (possible bias in that long latency responses might be either smaller or larger than those falling in the predetermined range), and latency (a restriction imposed by an arbitrary range). It would appear that it is the E's responsibility to reduce these distortions as much as possible by defining a response interval which accommodates S's behavior. In any event, the problem is neither specific to amplitude measurement nor relevant to whether or not amplitude, per se, is a useful measure.

Kimmel states that the definition of a nonresponse (or what gets entered as a zero in a magnitude analysis) is arbitrary. This arbitrariness is a problem in GSR conditioning, since different Ss are run at different sensitivity settings on the recording equipment. The result is that a nonresponse for one S may, in fact, be recorded as a response for another S. For example, for one S a minimum response of 1000 ohms decrease in resistance is required before a response is recorded, while for other Ss in the same study a minimum of only 100 ohms may suffice. This discrepancy has to do with the fact that UCRs and other observed responses in some Ss are quite large, while in others they are small, and that the sensitivity setting of the equipment is geared accordingly. Ideally, the solution would be to employ the same minimum for all Ss, but this is difficult with the equipment currently available. Until the recording problem is solved, we may have to continue to resort to different sensitivity settings (and, therefore, to different definitions of a response) for different Ss. However, this problem produces distortions in more than the definition of response probability. In magnitude measurement, the zero entry is weighted more heavily (i.e., is numerically further from the next interval of measurement) as the sensitivity of recording decreases. For latency measurement, there is simply a loss of latencies for those responses unobserved with the less sensitive recording. Similarly, average amplitude can be biased upward. Again, the problem, though real, is neither specific to amplitude measurement nor relevant to whether or not amplitude, per se, is a useful measure.

Kimmel also raises questions concerning the distribution of

amplitudes. His position seems to be that the eyelid response is described nicely by a dichotomous classification (i.e., large response vs little or no response), but that the eyelid response does not provide a model for the GSR. He states: "extension of this logic to the GSR ... ignores the fact that the distribution of the sizes of the GSR is not of this bimodal type. The GSR is a *continuously* variable quantity and a very, very small response is, merely, one that is slightly larger than a zero response."

There are several matters here. First, the eyelid response is not a simple dichotomous response. It varies substantially across trials, between treatments and between Ss in latency (Ebel & Prokasy, 1963), amplitude (Humphreys, 1943), volume (Pennypacker, 1964) and recruitment (Hartman, Grant, & Ross, 1960). The point is that all of the information contained in the eyelid response is not included in response probability, even though that has become the most frequently employed measure. It might be noted that a case was once made for the use of magnitude in eyelid conditioning on the grounds that it is a more reliable measure than either amplitude or probability alone (Humphreys, 1943). In any event, regardless of how the eyelid response is characterized, it is not particularly relevant to whether or not GSR amplitude is a useful measure.

Second, it would be useful to examine some GSR amplitude distributions, and these are included in Figs. 1 and 2. They come from a study involving a two by two design. One independent variable was CS intensity and the other was sensitization vs conditioning operations (Prokasy & Ebel, 1967). These are distributions of what Prokasy and Ebel called "first responses," the first response to occur within the

Fig. 1. Distributions of GSR amplitudes (in ohms) for two sensitization control groups. The ordinate provides the proportion of trials on which a particular amplitude occurred. The specific resistance change can be obtained by multiplying the category number by the inset number in each subfigure. The "0" category includes all trials on which a resistance change did not meet criterion, the value of the inset number in ohms.

Fig. 2. Distributions of GSR amplitudes (in ohms) for two conditioning groups. The ordinate provides the proportion of trials on which a particular amplitude occurred. The specific resistance change can be obtained by multiplying the category number by the inset number in each subfigure. The "0" category includes all trials on which a resistance change did not meet criterion, the value of the inset number in ohms.

interval of 1.35-4.95 sec after CS onset. As noted, all but 2% of the responses occurred within 4 sec after CS onset. Figure 1 presents the amplitude distributions at three sensitivity levels for the two CS intensities in the sensitization control groups. Figure 2 provides amplitude distributions for the two conditioning groups. These distributions include all responses made on the last 31 conditioning (sensitization) trials at the different sensitivity levels. The data of a small number of Ss are not included. These were Ss whose responses were recorded at more sensitive levels, but there weren't enough of them to generate stable frequency distributions. The "0" category includes those trials on which no recorded response occurred.

Whether or not the distribution is bimodal seems to be a joint function of treatment condition and sensitivity level. The sensitization treatment obviously results in unimodal distributions, either with or without the "0" category. The distributions are also unimodal at all sensitivity levels in both conditioning groups, provided that the "0" category is not included. Inclusion of that category converts three of those distributions to bimodality, with one of the modes being the "0" category.

While the conditioning and sensitization distributions overlap to some extent, it is plausible to distinguish between them by assigning differential weights to each of three sources of influence: noise level, orienting responses, and responses contingent on CS-UCS pairing. The noise level is independent of external signal control, and a response may or may not occur in the defined CR interval. The orienting reaction is tied to external stimuli, and increases in signal intensity result in increases in response frequency uniformly across sensitization and conditioning operations. Responses contingent on CS-UCS

pairing (i.e., the CR) are present only in the conditioning distributions. They are larger than those attributable to the other sources, and their occurrence results in an increase in both amplitude and probability.

The preceding analysis is incomplete and, indeed, may be wrong. However, it does illustrate how one can make use of both the bimodality of the GSR distribution and the separation of magnitude into its probability and amplitude components. The exclusive use of magnitude would not permit such an analysis, which serves to emphasize that an examination of alternative measures is itself a tool to be employed in the understanding of behavior modification.

It might be argued that the dominant characteristic of the sensitization groups, the "0" classification, is merely the accumulation of many small responses, and that to place them in a category of nonresponse without averaging them (at a zero value) into the magnitude measure is a distortion in that it doesn't recognize the existence of small responses. I do not find this a compelling argument against the use of amplitude, if for no other reason than that argument can be brought to bear on virtually any analog system. However, magnitude need not be accepted as the only index in GSR conditioning because the GSR is, by its very definition, probabilistic. That we can define the onset and peak of a response tells us that there are points in time when an increase in conductance does not occur. It is a system which can be defined, in part, in terms of the likelihood of a response in some unit of time. If the time unit is large, or if the minimum response criterion is very small, response probability will be high. Conversely, with smaller time units, or larger criteria, response probability will be lower. It in no way violates useful precedents in measurement to extend the idea of probability to the time interval during which a conditioned GSR is to be recorded. That a response can be "very, very small" does not change this; it is, after all, whether that response exceeds the criterion for defining a CR, and how it relates to overall baseline activity, that is important.

In this connection, Kimmel states: "The use of the amplitude measure assumes that zero response represents an entirely different kind of event (or nonevent)." This, it seems to me, is what the separation of magnitude into its amplitude and probability components is all about. The empirical question to be resolved is whether or not orderly information can be obtained by separately analyzing the data for the size of a response, given that one exceeds the criterion, and response probability. To my knowledge there is no strong a priori rationale for assuming that the failure to make an observable response is actually a response, though that assumption appears to be implicit in a magnitude analysis. To the contrary, it is less plausible to assume that GSR probability is 1.0 independent of the duration of the measurement interval, for reasons described in the preceding paragraph.

In concluding his paper, Kimmel notes that "a hoped for empirical resolution of the question may actually be

impossible" but that "the question may be resolved a priori as soon as the experimenter applies the mensurational logic of dichotomous responses to the continuous measurement situation." It is not entirely clear to me what the statements mean, but it does appear that Kimmel is saying that whether or not GSR amplitude measurement is useful may not be amenable to empirical analysis. If that is the case, I would argue, to the contrary, that in the long run it is the only possible solution. Historically the use of any psychological measure has been justified by the orderliness obtained when it is employed as an index of behavior. This orderliness may emerge from single experiments, or from comparisons across experiments, or it may emerge as a consequence of a theory or model which implies a specific measure. The question is not one of whether GSR amplitude is an intrinsically superior or inferior measure; rather, it is which of the many possible ways to code the GSR is the most useful for empirical and theoretical purposes. It may be that magnitude is sufficiently good as an index to preclude the need for amplitude and probability analyses. However, it is clear that unless magnitude is partitioned into its probability and amplitude components, no one will ever know whether the latter two measures are useful. In any event, as long as magnitude, as well as amplitude and probability, analyses are reported, there is no loss of information and it remains possible that we will, as a result, educate ourselves somewhat more concerning those mechanisms which underlie GSR conditioning.

REFERENCES

- EBEL, H. C., & PROKASY, W. F. Classical eyelid conditioning as a function of sustained and shifted interstimulus intervals. *Journal of Experimental Psychology*, 1963, 65, 52-58.
- GORMEZANO, I. Classical conditioning. In J. B. Sidowski (Ed.), *Experimental methods and instrumentation in psychology*. New York: McGraw-Hill, 1966.
- HARTMAN, T. F., GRANT, D. A., & ROSS, L. E. An investigation of the latency of "instructed voluntary" eyelid responses. *Psychological Reports*, 1960, 7, 305-311.
- HUMPHREYS, L. G. Measures of strength of conditioned eyelid responses. *Journal of General Psychology*, 1943, 29, 101-111.
- KIMMEL, H. D. GSR amplitude instead of GSR magnitude: Caveat emptor! *Behavior Research Methods & Instrumentation*, in press.
- MORROW, M. C., & KEOUGH, T. E., III. GSR conditioning with long interstimulus intervals. *Journal of Experimental Psychology*, 1968, 77, 460-467.
- PENNYPACKER, H. S. Measurement of the conditioned eyelid reflex. *Science*, 1964, 144, 1248-1249.
- PROKASY, W. F., & EBEL, H. C. GSR conditioning and sensitization as a function of intertrial interval. *Journal of Experimental Psychology*, 1964, 67, 113-119.
- PROKASY, W. F., & EBEL, H. C. Three components of the classically conditioned GSR in human subjects. *Journal of Experimental Psychology*, 1967, 73, 247-256.
- PROKASY, W. F., FAWCETT, J. T., & HALL, J. F. Recruitment, latency, magnitude, and amplitude of the GSR as a function of interstimulus interval. *Journal of Experimental Psychology*, 1962, 64, 513-518.

NOTE

1. Supported by funds from NIMH Grant MH15353-01.