*Journal of Experimental Psychology*
1974, Vol. 103, No. 1, 171–173

# SHORT REPORTS

## COMPENSATORY EFFECTS IN MORAL JUDGMENT: TWO RIGHTS DON'T MAKE UP FOR A WRONG[1]

DWIGHT R. RISKEY[2]
*University of Califironia, Los Angeles*

MICHAEL H. BIRNBAUM
*Kansas State University*

Ratings of the morality of persons described as having committed moral and immoral behaviors indicate that good deeds do not make up for bad ones. The overall goodness of a person is determined mostly by his worst bad deed, with good deeds having lesser influence. Addition of moral deeds does improve ratings of sets containing low-valued items, but, consistent with previous research, this compensation appears to be limited. Data suggest that performance of very immoral deeds limits the highest level of morality a person can achieve. The value of that limit appears to depend upon both the immorality of the bad deeds and the virtue of the good ones.

What rules of judgment describe how one forms an impression of the overall morality of another individual? Birnbaum (1972, 1973) found that ratings of the overall morality of a person described as having carried out several actions are not consistent with simple additive or averaging formulations. A person's morality does not appear to be the sum or average of the moralities of his actions. Instead, the individual's most immoral deed appears to have special significance.

With addition of highly moral items to sets containing an immoral item, the judgment of morality appeared to be limited to an unfavorable rating (Birnbaum, 1973). The present study replicates and extends the investigation of this effect to explore two simple possibilities: (*a*) that the upper limit is determined entirely by the most immoral item, and (*b*) that the upper limit depends upon the goodness of the deeds added to the immoral item.

*Method.* The *S*s were instructed to

> read each set of actions and then judge how "good" or "bad" it would be to carry out *all* of the actions .... In other words ... how morally "commendable" or "reprehensible" a person would be who carried out *all* of the actions.

Ratings were made on a 17-point scale ranging from −8 (very very bad) to +8 (very very good) in which zero was designated as neither good nor bad.

The *S*s were 50 University of California, Los Angeles, undergraduates who were fulfilling a class requirement in introductory psychology.

The 44 items used to construct the sets of behaviors were selected from a pool of 300 items that were previously rated by 101 *S*s on a 9-point scale (see Birnbaum, 1973). Typical items for the four levels are as follows: highly immoral (L), *Secretly spiking a party's potato chips with a dangerous drug;* moderately immoral (M−), *Cheating on an examination by copying from another student;* moderately moral (M+), *Fixing a friend's car for free;* and highly moral (H), *Rescuing a family from a burning house.*

The design was composed of 37 sets of two types: (*a*) Homogeneous sets of all M+ or all H items were constructed of 2, 3, 5, 7, 9, and 11 items. In addition, 1 set of 11 L items was included. (*b*) Heterogeneous sets were constructed according to a 6 × 2 × 2 (Number × High × Low) factorial design, in which 0, 1, 3, 5, 7, or 9 items of either M+ or H morality level were added to 2 items of either L or M− value. Except for the set of 11 L items, which received a mean judgment of −7.96, the design is represented in Figure 1.

There were four replicates of the entire design, each replicate containing different M− and L items. The M+ and H items were the same for each replicate although their printed order within sets was randomized, as were the items in the single set of 11 L items. In heterogeneous sets, the L or M− items were printed first.[3]

Booklets were constructed containing two full replicates. Each replicate included all 37-set types printed in random order on several pages. Pages were then assembled in all possible permutations. The cover page of each booklet contained the written instructions and response scale. The *S*s were required to skim the booklet before beginning

[1] Computing funds were provided by Campus Computing Network, University of California, Los Angeles. The second author received support from a National Institute of Mental Health Postdoctoral Fellowship at the Center for Human Information Processing, University of California, San Diego, which provided additional support through National Institute of Mental Health Grant MH-15828. The authors thank Allen Parducci and Norman H. Anderson for their helpful comments on the manuscript.

[2] Requests for reprints should be sent to either Dwight R. Riskey, Department of Psychology, University of California, Los Angeles, Los Angeles, California 90024, or Michael H. Birnbaum, who is now at the Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, Illinois 61820.

[3] A subsequent study tested the effect of the order in which simultaneously available items are printed. There were 46 *S*s who judged sets of 11 items in which 9 of the items were either M+ or H in value and 2 items were either L, M−, M+, or H in value. The location of the 2 L or 2 M− items was either at the beginning, middle, or end of the list. Results were similar to those reported in Figure 1. For example, the sets of 2L9M+ items received mean judgments of −4.0, −3.9, and −4.2 when the 2L items were at the beginning, middle, and end of the list, respectively. The printed order had negligible effects.

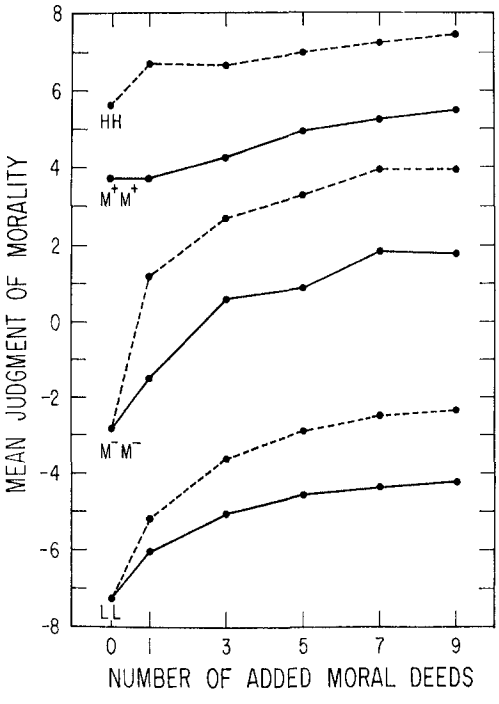


FIGURE 1. Mean judgments of morality as a function of the number of added moral deeds. (Each point represents the mean judgment of a different person as described by sets of actions varying in level of morality. Dashed curves show the effects of adding highly moral, H, deeds, and solid curves show the effects of adding moderately moral, M+, deeds; L = highly immoral, and M − = moderately immoral.)

to actually record their ratings. Half of the *S*s used either booklet.

*Results and discussion.* Each point in Figure 1 represents the mean judgment of morality of a different hypothetical person, described by a set of items. For example, the point labeled LL refers to the mean judgment of a person who has committed two highly immoral (L) deeds. The dashed curves show the effect of adding highly moral (H) deeds; the solid curves show the effect of adding moderately moral (M+) deeds. For example, starting at the point labeled LL, the points connected by the dashed curve show the morality of persons who have done 2L deeds and 1H deed, 2L and 3H deeds, etc.

According to the averaging model of impression formation (Anderson, 1971), the three dashed curves should converge, approaching an asymptote at a common limit (the value of H). Similarly, the three solid curves should converge to a common asymptote at the value of M+. Perhaps, with enough H deeds the curves would eventually converge, but as far as these data go, they show no indication of the convergence predicted by the averaging model. Contrary to the averaging model, the four bottom curves all appear to approach different asymptotes.

The lowest solid curve in the figure suggests that with the performance of 2L deeds, the addition of M+ deeds does not raise the overall morality above a judgment of −4, or "immoral." The dashed curve directly above it suggests that no number of H deeds would make the overall rating favorable. Having committed very bad deeds, one cannot expect to redeem himself by good actions.

The picture is not so bleak for peccadillos (M−). Having committed 2M− deeds, it is possible to be rated "good" by doing 1H or 3M+ deeds. However, inconsistent with the compensatory prediction of the general averaging model, the curves do not appear to approach a common asymptote with the curves for homogeneous moral items.

The two top curves, representing sets of homogeneous scale value, appear nearly linear and show little indication of an asymptote. The residuals from linearity for the dashed and solid homogeneous sets were nonsignificant, *F*s (4, 396) = 2.10 and 1.60, respectively, giving no evidence for an asymptote. Of course, with a greater number of items (more than 11) an asymptote might eventually be reached. Additionally, these end judgments may be distorted by a slight nonlinearity in the rating scale.

One possibility suggested by Birnbaum (1973) is that the upper bound would be completely determined by the most unfavorable deed. If this had been the case, the two lower curves would have converged to a common limit since they include the same immoral items. Similarly, the next two curves would have converged to another common asymptote.

Another unsupported possibility is that the level of the asymptote would be completely determined by the goodness of the deeds added to the immoral items. Thus, the lower dashed curves in Figure 1 would have approached the same asymptote. Similarly, the two lower solid curves would have converged to a common limit.

Instead of either of these simple possibilities, however, it appears that the limit depends both on the value of the worst deed as well as the value of the added good deeds. Perhaps the limit can be conceived of as an internal stimulus that is averaged in with the other information.

The present data indicate that ratings of morality show a peculiar form of compensatory effect. The findings that variation in the value and number of favorable deeds can raise the morality of low-valued sets may be called compensatory effects. However, the apparent finding of asymptotes suggests that the low-valued deeds, in combination with the added moral items, set some kind of upper bound. Given a person has done evil, an infinite number of good deeds may not produce a favorable overall impression. In this sense, morality judgment might be called "noncompensatory." As Ezekiel (*The New American Bible*, 3:20) has it, "If a virtuous man turns away from virtue and does wrong when I place a stumbling block before him, he shall die. He shall die for his sin, and his virtuous deeds shall not be remembered ...."

REFERENCES

Anderson, N. H. Integration theory and attitude change. *Psychological Review*, 1971, **78**, 171–206.
Birnbaum, M. H. Morality judgments: Tests of an averaging model. *Journal of Experimental Psychology*, 1972, **93**, 35–42.
Birnbaum, M. H. Morality judgment: Test of an averaging model with differential weights. *Journal of Experimental Psychology*, 1973, **99**, 395–399.
*The New American Bible*. New York: Catholic Publishers, 1971.

(Received August 15, 1973)

*Journal of Experimental Psychology*
1974, Vol. 103, No. 1, 173–174

# EFFECT OF SIGNAL FREQUENCY ON AUDITORY AUTOKINESIS[1]

G. RUSSELL[2] AND W. G. NOBLE
*University of New England, Armidale, New South Wales, Australia*

The magnitude of auditory autokinetic movement varied systematically with change in signal frequency from .5–10 kHz. Other response measures showed no such systematic trend. The latter finding confirms previous research results. Data allow an explanation of the phenomenon more in terms of interaural differences in adaptation than in terms of minimum audible angle.

Recent studies of apparent movement of an unseen, stationary sound source have failed to find any systematic relationship between acoustical frequency and either number of movements, mean movement duration, or movement latency. Perrott and Fobes (1971) advanced—and have since rejected (Fobes & Perrott, 1972)—a theory of auditory autokinesis based on auditory adaptation and localization. They stated that

> To account for auditory autokinesis, one need only to postulate that constant stimulation does not produce a constant level of activity in the peripheral auditory system. For the present model, sensory adaptation would provide the necessary change in neuronal activity to an invariant auditory stimulus. Any interaural differences in the adaptation function would provide "apparent" changes in interaural stimulation, which, in turn, could be coded as movement [Perrott & Fobes, 1971, p. 175].

For this theory to be supported, a frequency dependency of auditory autokinesis would need to be shown, and one which paralleled the frequency dependency of localization shift due to interaural intensity change (Mills, 1960). The lack of such a finding led Fobes and Perrott (1972) to reject the theory. But, failure to find a relationship between signal frequency and autokinetic movement could have been due to the very response measures used by these researchers. It is unclear why temporal features of the phenomenon were used as dependent variables. It would seem more logical to examine the spatial features (distance and direction) of what is after all a spatial phenomenon.

In a study by Bernadin and Gruber (1957), *Ss* reported that a tone of 5,600 Hz. apparently moved over a greater range than tones of either 56 or 560 Hz. Such a frequency effect is consonant with a prediction from the work of Mills (1960). Mills plotted (for varying signal frequency) the minima of interaural intensity differences (IIDs) that are required for listeners to lateralize a centered dichotic stimulus. He found that the minimum intensity difference was greatest at 1 kHz., decreased somewhat at frequencies below 1 kHz., and decreased even more at frequencies above 1 kHz., reaching the smallest value at 10 kHz.

It is known that adaptation is more or less constant across frequencies from 1 to 8 kHz. (Jerger, 1957), and that listeners are able to use IIDs to localize sounds. Thus, it would be expected that if interaural differences in adaptation occur, the resulting trend in magnitude of autokinetic movement at different frequencies should be inversely related to the trend for IIDs. By relying on response measures other than magnitude of apparent movement, Fobes and Perrott (1972) would not readily have discovered whether such a relationship exists.

Fobes and Perrott (1972) also argued that their failure to find a frequency effect is contrary to an alternative explanation of autokinesis—an explanation in terms of localization precision. They tested and subsequently rejected the hypothesis that the amount of autokinesis is inversely related to localization accuracy. But again, the response measures used are inappropriate for a test of this explanation. It may be that listeners can report apparent movement of a stationary source as occurring only within the spatial limits of minimum audible angle (MAA) at the same acoustical frequency. Under such a hypothesis, auditory autokinesis would be conceived as an "error" phenomenon, allowed by the limits of uncertainty within the difference limen for audible angle. If it were the case that auditory autokinesis and MAA are related, then the magnitude of apparent movement of a stationary sound source should parallel over frequency the minima of discriminable change of position of an actually

[1] This research was partly supported by an internal research grant from the University of New England.
[2] Requests for reprints should be sent to G. Russell, who is now at the Department of Psychology, Macquarie University, North Ryde, New South Wales 2113, Australia.