Psychological Bulletin
2012, Vol. 138, No. 6, 1269–1273

0033-2909/12/$12.00 DOI: 10.1037/a0029509

REPLY

What's All the Noise? Differentiating Dimensions of Acoustic Stress and the Limits to Meta-Analysis: Reply to Smith (2012)

J. L. Szalma and P. A. Hancock
University of Central Florida

Smith (2012) has provided pertinent observations on our recently published meta-analytic review (Szalma & Hancock, 2011) of the effects of acoustic noise on performance. His main points are as follows: (a) our review excluded some areas of research; (b) there were conceptual problems with our moderator analyses; and (c) limitations to meta-analysis, in general, constrain the conclusions that can be drawn from our findings. In this response, we address these issues and, in so doing, account for Smith's concerns and consequently identify useful avenues for future work. Thus, we argue that (a) most of the areas of research described by Smith were outside the explicitly specified and so-stated limits of our meta-analysis; (b) the conceptual problems in our meta-analysis are common to all research on stress and performance; and (c) the cited limitations of meta-analysis have been well established previously. We therefore remain unchanged in our opinion that meta-analysis is a powerful tool for quantifying the effects of variables on outcome measures across collective studies. Although acknowledging certain inherent procedural limitations, we nevertheless agree that a full comprehension of noise effects on performance has yet to be exhaustively articulated.

Keywords: noise, stress, performance, meta-analysis

Noise is surely both an interesting and divisive issue. Indeed, this is the second occasion in this journal alone wherein a pointed comment has appeared in response to a target article on the topic (see Broadbent, 1978; Poulton, 1977, 1978). Smith (2012) has raised interesting concerns with our meta-analytic review of the effects of acoustic noise stress on human performance (Szalma & Hancock, 2011). These observations point to certain inherent boundary conditions to the process of meta-analysis itself. We thank Smith for this opportunity to elaborate on a number of the issues raised and to further extend our position.

One of the primary purposes for our published analysis (Szalma & Hancock, 2011) was to provide a seminal quantification of the categories of noise stress research to reveal the nature and dimensions of the important gaps that remain. We believe that Smith's (2012) piecemeal description of a number of different categories, tasks, and noise characteristics actually reflect the clear and systematic success in our meeting this stated goal. Thus, we appreciate that his commentary confirms our conclusions and our explicit observations regarding substantial gaps in the literature.

J. L. Szalma, Performance Research Laboratory, Department of Psychology, University of Central Florida; P. A. Hancock, MIT^2 Laboratory, Department of Psychology, University of Central Florida.

Correspondence concerning this article should be addressed to J. L. Szalma, Performance Research Laboratory, Department of Psychology, 4000 Central Florida Boulevard, Psychology Building 99, University of Central Florida, Orlando, FL 32816. E-mail: james.szalma@ucf.edu

The Scope of a Meta-Analysis

The purpose of a meta-analysis is to provide quantitative synthesis of the findings of a proscribed research literature, with studies selected based on a pre-determined set of criteria derived from theoretical, practical, and analytic concerns. Our own work therefore differs from each of the previous reviews noted by Smith (2012) in that the latter were ubiquitously qualitative summaries, whereas ours, in contrast, consisted of statistical estimation of the quantitative effects of noise stress on performance. That our review (Szalma & Hancock, 2011) reached conclusions consistent with those cited by Smith is gratifying, confirming the accuracy and utility of our work. In some instances, noise has no effect on performance (Kryter, 1970); in other cases, the effects are complex and subtle (Loeb, 1986); and in yet other cases, the effects depend directly on task type (Smith, 1989). In Szalma and Hancock (2011), we provided explicit and quantitative enumeration beyond these previous qualitative assessments, also cataloging the precise values of the absolute and relative magnitudes of these respective effects—and how they changed as a function of our identified moderator variables. Our review therefore extended well beyond all the previous qualitative accounts and reviews.

Boundary Conditions and Selection Limitations

One of the stated principal boundary conditions for our meta-analysis was the restriction to healthy adults. As such, we intentionally excluded all studies involving children and clinical populations. Furthermore, our concern, as we clearly stated, was with acoustic noise effects in which tasks were performed during actual exposures, a strategy we have adopted in companion meta-analyses on differing physical sources of stress (Conway, Szalma, & Hancock, 2007; Hancock, Ross, & Szalma, 2007). Thus, aftereffects of noise were not relevant to our review. The study of noise aftereffects per se (or, indeed, aftereffects of any source of stress exposure) is, of course, of great theoretical and practical importance (see, e.g., Harris, Hancock, & Harris, 2005), as indeed are consideration of noise effects in children and special populations. However, these lay beyond our current purview; as indicated in our inclusion criteria, they were never intended to be within the scope of our meta-analysis. The exclusion of children was partly for pragmatic reasons concerning the funding of our work, but also because there are potential developmental interaction effects that might provide yet another moderator regarding our main conclusions. Indeed, we join with the implicit admonition of Smith (2012) in recommending a future meta-analytic review of the specific noise effects on this important group (children).

Smith (2012) also correctly pointed out that there have been many field studies of noise effects and that we did not incorporate these studies in our review. However, again among our explicit criteria for inclusion was the requirement for an empirical examination of noise stress with an experimental manipulation and use of a control group. As in the case of aftereffects, therefore, nonexperimental studies were excluded. We confirm that field studies are most certainly important, but we intentionally excluded them in our work. We appreciate also that Smith pointed to papers presented at the International Commission on Biological Effects of Noise (ICBEN) conference, and it is an interesting coincidence that this meeting occurred in the same month as the formal publication of our article. One of the inherent problems in all meta-analytic research, of course, is that of obtaining unpublished findings given in places such as presentations, posters, and internal laboratory reports. This is a limitation common to all meta-analyses (Hunter & Schmidt, 2004).

Smith (2012) noted that meta-analytic approaches often do not encapsulate the variety of strategies that individuals use to perform tasks, and that such variation in strategy use is particularly important for noise research. We agree that strategy use is important, and indeed its variation is common to most if not all sources of stress and workload. It is not possible, however, to examine these meta-analytically without the empirical database that permits such moderator analyses. In other words, strategy use (or any variable of interest) can only be included in a meta-analysis if it can be unambiguously coded from each original report, and if there are a sufficient number of studies to provide stable estimates of central tendency and variability.

Smith (2012) asserted that in our review we did not reject studies based on methodological weaknesses. We adopted, and still advocate, an empirical approach to this issue. Whether methodological quality should be included as a criterion has indeed been controversial, and there are two general approaches toward a resolution: (a) include only the studies of the purported "best" quality or (b) include all studies and empirically evaluate whether the "poor" studies disproportionately influence analytic outcomes (Lipsey & Wilson, 2001). We used the latter approach, as it provides a quantitative basis for assessment. Our meta-analysis accounts for study quality to some degree by weighting effect sizes by the reciprocal of the variance. Poorly designed studies are more likely to have larger variances, resulting in a relatively weaker contribution to the overall weighted average effect size (Hedges & Olkin, 1985).

Analogous to the prior point, Smith (2012) pointed out that noise can have different effects on different stages of information processing. We find this line of inquiry interesting and clearly worthy of further empirical pursuit, but to evaluate this analytically from the extant literature would again require unambiguous coding criteria for categorizing tasks reported into their information processing stages. Smith correctly noted that intensity (I) of noise and the duration of exposure (T) may interact, and that the effects of noise depend on task type. We agree with both statements, but arguing that continuous loud noise exposure for a short duration is similar in effect to longer exposures with more moderate intensities implies that $I \times T =$ constant, which is unlikely. Although higher intensity but short duration noise may be *functionally* equivalent to moderate intensity exposures over longer time epochs, it does not mean that they are equivalent in their effects on information processing. It is also unlikely that the effects of noise (or any source of stress) on the performance of a complex task will be limited to a simple combination of the basic cognitive components of the task.

Noise Characteristics and Individual Differences

Many recent laboratory studies of acoustic distraction concern the effects of irrelevant speech, as Smith (2012) correctly noted. Indeed, we emphasized this dimension and its importance in our original article. However, our intent was to examine the effects of noise as a source of *stress*, not of distraction per se. Although distraction may well contribute to change in performance, we remain firm in the conclusion that our results are consistent with a broader, resource-depletion account of acoustic noise effects.

Smith (2012) noted that we did not explicitly consider individual differences in noise effects on performance. Again, we strongly agree that consideration of individual differences is a crucial aspect of research on stress and performance, and we have pursued this issue ourselves (see Hancock, Hancock, & Warm, 2009; Hancock & Warm, 1989; Szalma, 2008, 2009). However, consideration of such variables was outside the scope of our meta-analysis. One cannot analyze individual differences in effect sizes at a single participant level because the variability computed in a meta-analysis is across *studies* and not across individual participants. The only way to evaluate individual differences in a meta-analysis would be to examine studies in which individual difference variables were measured, an effort we strongly advocate. Perennially, the tension between experimental conditions and person characteristics has been a central concern for all of psychology (Cronbach, 1957, 1975).

Stress Theory, Stress Interactions, and Noise

Smith (2012) noted that extant theoretical models of stress and performance fail to account for the range of effects of noise. We are rather puzzled by this comment, as the data from our analyses indicated strong support for a resource theory account of noise effects. In addition, our discussion of arousal was not based on views from the neuroscience literature that do not provide explicit theories of noise effects, but rather from the perspective concerning first the unitary and then the fractionated arousal mechanisms applied to stress and emotion generally (Easterbrook, 1959; Hebb, 1955) and to noise specifically (Broadbent, 1978; Poulton, 1979). However, Smith was correct in pointing out that a key issue remains in defining "noise" itself. This is also true of stress in general, because each is essentially a relational concept. That is, the stress associated with a stimulus depends not simply on its physical characteristics but also its meaning to the individual performer (cf. Hancock & Warm, 1989; Lazarus, 1999; Lazarus & Folkman, 1984).

Smith (2012) protested that one of the strongest forms of support for the arousal explanation of noise effects relates to the apparent improvement in performance of individuals who are, for various reasons, assumed to be in a state of low arousal. We have two forms of response to this important observation. The first relies on the persistence of our protestation that we surveyed the capacities of healthy and alert individuals. Thus, reference to performers who were experiencing a common cold (Smith, Thomas, & Brockman, 1993), or who were suffering a hangover (McKinney & Coyle, 2007), or who had been deprived of sleep for more than 50 consecutive hours (Corcoran, 1962) falls beyond the purview of the boundaries of our meta-analytic inclusion criteria. However, in part, this is an insufficient answer to Smith's observations. As a result, we briefly examine some of the cited evidentiary basis for Smith's assertions.

It is clear that Corcoran's (1962) finding, at least for his first experiment, derives from a highly doubtful comparison procedure using subtraction methods across between-subject groups of very limited numbers of participants. Indeed, Corcoran's own concluding observation on his second experiment that "there was a suggestion in the NS [no sleep] conditions that gaps [in the performance] measure were somewhat reduced by noise . . . [and] errors showed no effect of either loss of sleep or noise" (p. 180) provides confirmation of this doubt and therefore little confidence in the reliability of such an influence. Further, Blake's (1971) experiment concerning noise is also highly suspect. He reported that in a quiet condition there was approximately a 9% increase in performance output from 8:00 a.m. until 10:30 a.m. on a letter cancellation task. However, accompanying this marginal output increase is more than an order of magnitude decrease in error rate during the same interval. The comparable numbers for production in noisy condition was a 3.5% increase compared to again an almost order of magnitude decrease in the level of error. Why a letter cancellation task yields such an enormous change in error after only 150 min of morning work is not apparent, yet one explanation might be found in a tragic footnote to this particular work. Sadly, the chapter was published after the author's untimely death, and the sections that were compiled by colleagues were explicitly noted. This relevant Experiment 10 was specifically among the latter compilations. Thus, it may be possible that these data were not meant for publication before further vetting. Our ultimate response concerning Smith's (2012) comment on low arousal is, of course, that there remains the unenviable problem of trying to provide independent specification of the overall state of arousal anyway, which must necessarily be verified by other methods than assertion alone.

Smith (2012) noted that "studies of the combined effects of noise and other factors rarely consider type of noise or type of task" (p. 1265). This is true, and we view our work as providing an important numeric resolution to this issue by first identifying noise effects as a function of noise and task type, and second by identifying gaps in the empirical literature that should be addressed to enhance our understanding of moderating contextual factors. Beyond factors already discussed, we also suggest that consideration of the relation of noise to differential effects on the two arousal systems (energetic and tense; Thayer, 1978, 1989) has been neglected to date (although see Helton, Matthews, & Warm, 2009) as well as in stress research in general (for a noteworthy exception, see Matthews et al., 2002). Of course, for highly pragmatic real-world assessments, there remain a lamentably small number of experimental studies in which the full spectrum of operational stresses (i.e., heat, noise, vibration, etc.) has been evaluated together (Hancock & Szalma, 2008).

The Limitations of Meta-Analysis

Regarding the comments about the general limitations of meta-analyses, it is correct to say that our meta-analysis did not explicitly consider the problems of asymmetric transfer between noise and quiet conditions. Consistent with our previous synthetic efforts (Conway et al., 2007; Hancock et al., 2007) and consistent with analyses of dimensions such as after-effects, examining asymmetric transfer effects would require the gathering and quantifying of relevant methodological specifications. However, based on the interactions between task type and noise characteristics that we found in our work, it is likely that transfer itself would exhibit further embedded interactive effects. Distillation of these respective influences would therefore require at least a third-order hierarchical evaluation, and it remains very unlikely, if not actually improbable, that the current empirical database could support such analyses.

Although collapsing across outcome measures can lead to misinterpretations, this is only true if, in fact, the different outcomes (or, more generally, the different levels of any moderator variable) reflect different phenomena. Meta-analytic techniques allow for relevant evaluation by computing the variability associated with the weighted average effect size and by the use of hierarchal moderator analysis to determine whether the effect sizes for the different levels of the variable differ substantially from one another. Hence, the misinterpretation of effect sizes collapsed across categories of a variable is of the same order as interpreting a main effect in an analysis of variance (ANOVA) in the presence of a significant interaction (Hunter & Schmidt, 2004).

The Logic of Hierarchical Meta-Analysis and the Apples and Oranges Problem

It is common practice to evaluate the effect of hypothesized moderator variables in meta-analysis. In our review (Szalma & Hancock, 2011), these included both task and noise characteristics. Separate examination of moderator variables is akin to tests of main effects in analysis of variance. However, such analyses can mask the presence of interactions; in meta-analysis, this issue can be addressed by hierarchical analyses in which two variables are examined together. For instance, we examined noise schedule (intermittent vs. continuous) within each level of noise type. Comparison of these effect size magnitudes (and of the respective confidence intervals) identifies interactive effects between variables (for a more detailed discussion, see Hunter & Schmidt, 2004). A limitation is that, as in the case of complex interactions in ANOVA, the number of data points used to compute "cell means" (i.e., the average effect size for each combination of two or more moderators) decreases with increasing complexity of interactions. Indeed, in our analyses, "three-way" moderator effects could not be reliably estimated in most instances precisely because of this constraint. However, we contend that even these identified cases have an important value precisely because they clearly show the gaps in empirical research that need to be addressed.

Smith (2012) also noted that our analyses were performed at a gross level and that a finer-grained treatment of task types is warranted. We generally agree, but we have to point out that meta-analysis derives trends but does not impose them. Even with the present level of our analyses, we were constrained by space limitations to the analysis of only two-level moderators (three-level moderator analyses were included as an online supplementary document: http://dx.doi.org/10.1037/a0023987.supp). The gross level of analysis was thus necessitated by the need for a sufficient number of effect sizes to derive stable estimates and the constrained magnitude of the review itself. Some relatively complex moderator analyses could be computed because there were sufficient numbers of effect sizes. In other cases, the number of studies in a category directly limited our hierarchical analyses. As Hunter and Schmidt (2004) noted, such findings indicate that "Firmer conclusions must await the accumulation of a larger number of studies" (p. 426).

An additional problem is deciding how levels of a moderator variable are to be categorized. This complicates any fine-grained analyses, as the number of potential categories of "cognitive tasks" could quickly proliferate to a number that makes meta-analysis impractical. An old issue in meta-analysis is that of the "apples and oranges problem," in other words, combining effect sizes from studies that examined different phenomena (Cooper & Hedges, 1994; Glass, McGaw, & Smith, 1981). The issue is determining what level of generalization is appropriate to the scientific questions at hand (Cooper & Hedges, 1994; Glass et al., 1981). However, it has been noted by several researchers that mixing apples and oranges is supportably appropriate if one wishes to understand fruit—that is, that mixing studies using different tasks is appropriate if the goal is to generalize to higher order characteristics (Glass, 1978; Rosenthal, 1990).

Smith (2012) noted that many noise studies involve several outcomes and that to analyze these may lead to misinterpretation. This relates to the aforementioned "apples and oranges" problem, but it is also one that we addressed in our meta-analysis via our selection criteria and the analytic method itself. First, we only included performance outcomes such that other outcome measures (e.g., subjective response) were explicitly omitted. Second, the methods of meta-analysis permit an evaluation of the appropriateness of combining effect sizes. By estimating the sampling error bias and subtracting it from the observed variance, one obtains an indication of the amount of residual variance, that is, the variance remaining after accounting for the variability due to sampling error. Instances in which there is substantial residual variance (e.g., cases in which the residual variances is greater than 25%; Hunter & Schmidt, 2004) indicate heterogeneity of effect sizes. Stated another way, a large residual variance suggests that the average effect size is not an accurate estimate of central tendency because there are moderator variables (or levels of a variable) not accounted for in that particular analysis. This is analogous to a main effect in an ANOVA being conditionalized by the presence of an interaction. In our meta-analysis, there were substantial numbers of cases in which the residual variance was large relative to the sampling error variance, indicating the presence of complex moderator effects (e.g., three- or four-way "interactions"), and we noted and reported these.

Conclusions

In summary, we agree with Smith (2012) on a number of points. Yet, most of his points were outside the scope and limits of our study, which we explicitly specified in our inclusion criteria. Limitations of meta-analytic techniques constrain the exploration of potential moderator variables and combinations of variables; a chief contributor to these limitations is the dearth of relevant empirical studies. We still believe that our meta-analysis identified many (although certainly not all) of these empirical gaps. We hope that one principal contribution of our review, as well as the commentary by Smith and this reply, will be to inspire such theoretically based empirical efforts to address these important remaining issues.

References

Blake, M. J. F. (1971). Temperament and time of day. In W. P. Colquhoun (Ed.), *Biological rhythms and human performance* (pp. 109–148). London, England: Academic Press.

Broadbent, D. E. (1978). The current state of noise research: Reply to Poulton. *Psychological Bulletin, 85,* 1052–1067. doi:10.1037/0033-2909.85.5.1052

Conway, G. E., Szalma, J. L., & Hancock, P. A. (2007). A quantitative meta-analytic examination of whole-body vibration effects on human performance. *Ergonomics, 50,* 228–245. doi:10.1080/00140130600980888

Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis.* New York, NY: Russell Sage Foundation.

Corcoran, D. W. J. (1962). Noise and loss of sleep. *The Quarterly Journal of Experimental Psychology, 14,* 178–182. doi:10.1080/17470216208416533

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671–684. doi:10.1037/h0043943

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30,* 116–127. doi:10.1037/h0076829

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review, 66,* 183–201. doi:0.1037/h0047707

Glass, G. V. (1978). Integrating findings: The meta-analysis of research. *Review of Research in Education, 5,* 351–379.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Newbury Park, CA: Sage.

Hancock, P. A., Hancock, G. M., & Warm, J. S. (2009). Individuation: The $N = 1$ revolution. *Theoretical Issues in Ergonomics Science, 10,* 481–488. doi:10.1080/14639220903106387

Hancock, P. A., Ross, J. M., & Szalma, J. L. (2007). A meta-analysis of performance response under thermal stressors. *Human Factors, 49,* 851–877. doi:10.1518/001872007X230226

Hancock, P. A., & Szalma, J. L. (2008). Stress and performance. In P. A. Hancock & J. L. Szalma (Eds.), *Performance under stress* (pp. 1–18). Aldershot, England: Ashgate.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31,* 519–537.

Harris, W. C., Hancock, P. A., & Harris, S. C. (2005). Information processing changes following extended stress. *Military Psychology, 17,* 115–128. doi:10.1207/s15327876mp1702_4

Hebb, D. O. (1955). Drives and the C. N. S. (conceptual nervous system). *Psychological Review, 62,* 243–254. doi:10.1037/h0041823

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York, NY: Academic Press.

Helton, W. S., Matthews, G., & Warm, J. S. (2009). Stress state mediation between environmental variables and performance: The case of noise and vigilance. *Acta Psychologica, 130,* 204–213. doi:10.1016/j.actpsy.2008.12.006

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting for error and bias in research findings.* Thousand Oaks, CA: Sage.

Kryter, K. D. (1970). *The effects of noise on man.* New York, NY: Academic Press.

Lazarus, R. S. (1999). *Stress and emotion: A new synthesis.* New York, NY: Springer.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping.* New York, NY: Springer-Verlag.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Loeb, M. (1986). *Noise and human efficiency.* Chichester, England: Wiley.

Matthews, G., Campbell, S. E., Falconer, S., Joyner, L. A., Huggins, J., Gilliland, K., . . . Warm, J. S. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion, 2,* 315–340. doi:10.1037/1528-3542.2.4.315

McKinney, A., & Coyle, K. (2007). Next-day effects of alcohol and an additional stressor on memory and psychomotor performance. *Journal of Studies on Alcohol and Drugs, 68,* 446–454.

Poulton, E. C. (1977). Continuous intense noise masks auditory feedback and inner speech. *Psychological Bulletin, 84,* 977–1001. doi:10.1037/0033-2909.84.5.977

Poulton, E. C. (1978). A new look at the effects of noise: A rejoinder. *Psychological Bulletin, 85,* 1068–1079. doi:10.1037/0033-2909.85.5.1068

Poulton, E. C. (1979). Composite model for human performance in continuous noise. *Psychological Review, 86,* 361–375. doi:10.1037/0033-295X.86.4.361

Rosenthal, R. (1990). An evaluation of procedure and results. In K. W. Wachter & M. L. Straf (Eds.), *The future of meta-analysis* (pp. 123–133). New York, NY: Russell Sage Foundation.

Smith, A. (1989). A review of the effects of noise on human performance. *Scandinavian Journal of Psychology, 30,* 185–206. doi:10.1111/j.1467-9450.1989.tb01082.x

Smith, A. (2012). An update on noise and performance: Comment on Szalma and Hancock (2011). *Psychological Bulletin, 138,* 1262–1268. doi:10.1037/a0028867

Smith, A., Thomas, M., & Brockman, P. (1993). Noise, respiratory virus infections and performance. *Actes Inrets, 34,* 311–314.

Szalma, J. L. (2008). Individual differences in stress reaction. In P. A. Hancock & J. L. Szalma (Eds.), *Performance under stress* (pp. 323–357). Aldershot, England: Ashgate.

Szalma, J. L. (2009). Individual differences in human-technology interaction: Incorporating variation in human characteristics into human factors and ergonomics research and design. *Theoretical Issues in Ergonomics Science, 10,* 381–397. doi:10.1080/14639220902893613

Szalma, J. L., & Hancock, P. A. (2011). Noise effects on human performance: A meta-analytic synthesis. *Psychological Bulletin, 137,* 682–707. doi:10.1037/a0023987

Thayer, R. E. (1978). Toward a psychological theory of multidimensional activation (arousal). *Motivation and Emotion, 2,* 1–34. doi:10.1007/BF00992729

Thayer, R. E. (1989). *The biopsychology of mood and arousal.* New York, NY: Oxford University Press.

Received May 11, 2012
Revision received May 18, 2012
Accepted May 21, 2012 ■