Challenges in quantitative dream research

A commentary on Curtiss Hoffman's "Research articles in *Dreaming*: A review of the first 20 years" published in *Dreaming*, 23, 216-231

Michael Schredl

Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Germany

Summary. The paper of Curtiss Hoffman addresses important issues relevant to quantitative dream research. Although the paper is somewhat problematic, I hope the paper and this commentary will help to stimulate a fruitful discussion about the methodological standards in dream research. In my view the most important issues are: Formulating explicit hypotheses derived from theory, reporting sample characteristics (age, gender, and other study-relevant variables), application of adequate statistics (if necessary consult a statistician), and an obligatory discussion of methodological issues within the paper.

Keywords: Dream research, quantitative methods, statistical issues

Firstly, I want to express my admiration for Curtiss Hoffman for putting himself out there and writing an article about the methodological quality of the research papers published in *Dreaming* (Hoffman, 2013). It's no accident that the paper, pointing out weaknesses, was not written by a quantitative dream researcher; as one is reluctant to throw the first stone and criticize the work of fellow colleagues in the field. To my knowledge, Milton Kramer is the only one who had systematically looked at the methodological quality of dream studies. He carried out extensive literature searches regarding research papers on psychopathology and dreaming and analyzed the publications according to the following criteria: adequacy of the description of the patient samples, adequacy of the description of the control samples (if control samples were included), description of dream collection methods, adequacy of dream content analysis and statistical methods (Kramer, 2000; Kramer & Nuhic, 2007; Kramer & Roth, 1978). His conclusion (Kramer, 2010) is not very encouraging: "The small number of studies in many of the entities of interest and the general lack of scientific rigor continues to limit any potential value in the study of dreams (p. 381)." Recently, Stumbrys, Erlacher, Schädlich, and Schredl (2012) carried out a systematic review of lucid dreaming induction studies. In addition to summarize the findings, the methodological quality was evaluated by adopting a widely-used rating system (Downs & Black, 1998). This rating system encompasses 27 items regarding internal validity, and external validity, and how the data were reported. Two independent judges who applied the rating system showed high interrater reliability (kappa = .91). Their conclusion was "The methodological quality of the included studies was relatively low (p. 1456)."; compared to the standards of other fields of academic psychology. These previous studies clearly indicate that there is a lot to improve regarding the methodological quality of empirical dream research.

So, thank you Curt, for starting an interesting and hopefully beneficial discussion. However, some of Hoffman's ideas and conclusions are misleading and need clarification, especially the topics "theory-driven" research, sampling strategies, and application of statistical methods.

But let me start with an obvious but sometimes forgotten fact, namely that dream research, especially within academic psychology, is a very small field (see Milton Kramer's comment on dream studies in patients with mental disorders). A keyword search performed on September 4th, 2013 in PsycInf, the most often used database in academic psychology (developed from Psychological Abstracts (printed index of scientific articles that started in 1927)), yielded 7237 hits for all the dream-related keywords of the thesaurus (Dream analysis, Dream Content, Dream recall, Dreaming). Even though the number is quite high, I checked the keyword "short-term memory" which yielded 17,088 hits, i.e., the number of studies looking at a very specific cognitive function exceeds by far all dream-related paper published and indexed in the database. Being a small field has its advantages (e.g., less competition between the researchers because there are a lot of different topics still untouched) but also its disadvantages, especially if the researchers are not full-time researchers, e.g. undergoing full training in research methodology and statistics. Curtiss Hoffman, for example, mishandled the Chi-square test (see Figure 11; Hoffman, 2013) and would not pass the basic statistics course in psychology. You cannot interpret any findings with p values larger than .05 and say that these are moderate correlations or even weak ones. The terms small, moderate, and large are terms that are related to effect sizes, see Cohen (1988). Even though, the example of Hoffman's inadequate handling of statistical measures is an obvious one and understandable as he is working in a different field, I also have to admit that my statistical training which I received over 25 years ago is helpful in some ways but other research questions are so complicated that I could not solve these issues by myself. Fortunately, our institute where I am working has

Corresponding address:
Prof. Dr. Michael Schredl, Sleep laboratory, Central Institute of Mental Health, PO Box 122120, 68072 Mannheim, Germany.
Email: Michael.Schredl@zi-mannheim.de

a statistical department with mathematicians who can advise whether or not sophisticated methods are necessary for analyzing the data. In small research fields like dream research these networks and resources are not often available. This is a problem because some paradigms in dream research needs very sophisticated analysis methods. This topic will be elaborated below.

Let me first start with the basic issue of generating hypotheses. Being an academic, teaching research courses for quite a while, I am aware of the pitfalls of carrying out empirical studies. A relatively common one is that the hypothesis is justified by referring to a previous empirical study reporting a significant finding for that particular topic. This is not the correct way because hypotheses can only be derived from theories (most likely the previous study had some kind of theoretical background). And it is noteworthy, especially in view of the strong emphasis on quantitative methods in academic psychology, that theory building and formulating hypotheses is a creative endeavor; there are no fixed rules about what theories are interesting, or what kind of hypotheses are relevant to promote the field and so on. Unfortunately, psychology students learn very little about this creative process. A look at Curt's paper shows that his research questions are not adequately embedded in theory, especially theories within the field of philosophy of science. The relationship between sample size, including gender distributions, and the number of statistical tests with mentioning caveats cannot be established by theory (it might reflect – as Curt pointed out – the personal style of a particular researcher). There is no doubt that an empirical study should include a description of the sample, for instance, age means and standard deviation, gender distributions and, especially in the USA, ethnicity. Ethnicity, for example, is of minor importance in countries like Germany. But the variables you could include are numerous starting with social background, income, intelligence, handedness, childhood trauma and so on. The researcher must focus on what is necessary within the context of his/her own research. But the main issue about whether it is important to have random samples or self-selected or experimenter-selected samples is the research question itself. If you want to look at gender differences in dreaming, for example, it might be necessary to complement findings in student samples with surveys in the general population because you would expect that the shared waking life environment of students (studying the same topics, social life with fellow students) might affect gender differences in dream content compared to women and men living, for example, in traditional roles (women who are caring for children and men earning money for instance). On the other hand, if you look at mechanisms, for example, is dream recall related to 'openness to experience' personality dimension, you can use any sample you like because you postulate that this hypothesis is valid for all persons. To make that more clear, you would expect that every human being will experience pain if you pinch him/her with a needle. Having said this, statistical issues have to be considered. If you select only high dream recallers for your study looking into the relationship between dream recall and personality, you will have a restricted variance in your dependent variable which affects the validity of the statistical results – most often you find smaller correlations within samples where the variance is restricted. That would be similar if you only include persons that are very pain sensitive and not the whole range of persons with different pain

sensitivities. If you assume that the variance of the variables under study is limited by your sampling method, mentioning caveats about this seems appropriate. Just briefly, the major issue with sample size in statistics is of course the statistical power of your tests. If you have a small sample size, the measurement error variance is large and can cover up "real" relationships between your variables, i.e., you have to be very careful to interpret non-significant findings obtained in small samples.

Another issue addressed by Hoffman (2013) is the number of statistical tests applied within a study. While he is looking whether the number of statistical tests are related to caveats and the kind of conclusions, in the field there is a lot of debate whether it is necessary to apply techniques of adjusting the p-value, because the widely used cut-off of $p = 0.05$ would produce on average a significant finding for every 20 tests you perform. In my opinion, this discussion is not very helpful for researchers. I would like to emphasize that it is much more important to formulate specific hypothesis based on theory and differentiate between testing these a priori hypotheses in contrast to exploratory analyses. To illustrate this, if you carry out a lab study with 20 participants and put in a lot of work, you can report the test of the hypothesis, for example, that REM dream recall rates are higher than NREM dream recall rates but you might also report some findings you didn't expect (as an explorative finding).

The last issue I want to address in this commentary is the adequate application of statistical methods. In the last decades, statistical methods have become more and more sophisticated. For example, for analyzing published results meta-analytic methods have been developed – not often applied in dream research (Beaulieu-Prevost & Zadra, 2007; Schredl & Reinhard, 2008, 2011). The basic idea of these complex methods is that relationship between effect sizes or correlation coefficients and other variables can be weighted with sample sizes of the individual studies (including the appropriate algorithms for specific distribution characteristics of the studies' effect sizes). For testing his research questions, Hoffman (2013) should have applied meta-analytic methods as an appropriate approach. Another complex method, which is helpful in large samples looking at relationships between a large variety of variables, is structural equation modeling, e.g. Punamäki (1998), Schredl, Wittmann, Ciric, and Götz (2003), and Yu (2013). An issue that is addressed by Hoffman (2013) under the topic of duration is the number of dreams per participant. Most statistical tests require as prerequisite an independence of observations, in this case, dreams. This is normally the case if each

Table 1. Suggestions for improving empirical articles.

Suggestions
• Formulating explicit hypotheses derived from theory
• Report sample characteristics (age, gender, and other study-relevant variables)
• Adequate statistics (if necessary consult a statistician)
• Discussion of methodological issues

dream is provided by a different participant. But in case of longer studies, for example, diary studies or lab studies, the researcher usually obtains several dream reports per participant. In order to deal with this issue (repeated measurements) one has to apply mixed models if the number of dreams per participant is varying (e.g., Schredl et al., 2012). The most difficult methodological challenge I encountered so far was the analysis of a long dream series with binary outcome variables. The problem is the fact that the time series in dream research is never complete, i.e., due to lack of recall the time lags between one measurement point to the next is varying. My colleague form the statistics department was able to find a solution (Schredl & Reinhard, 2012) but even this algorithm was not working in all cases.

To summarize, a considerable number of research questions in dream research require sophisticated statistics; an endeavor which often can only be handled if specialists are consulted.

Conclusions

Again, I would like to thank Curtis Hoffman for publishing this paper and for starting a broad discussion among researchers about carrying out and publishing quantitative dream research. Despite the weaknesses of his paper, he pointed out how the publications in *Dreaming* can improve in the future. My suggestions are listed in Table 1. As I pointed out, one of the major issues is to formulate one or more hypotheses that are derived from a theory that is accepted (or at least discussed) in the field. That is an issue where the blind reviewer serving for *Dreaming* might help. The editorial staff might come into action if the researcher does not provide sufficient information about the sample: For instance, if some information is not available, the researcher should be encouraged to discuss this issue. The use of adequate statistics is of course a trivial suggestion but, as I pointed out above, there are some research questions that require sophisticated methods. In Germany, for example, researchers organize advanced statistics workshops (structural equation modeling or analyzing fMRI data) to help the researchers to cope with these challenges.

Finally, there is another point with which I disagree in Hoffman (2013): In my opinion, a discussion of methodological issues like sampling methods, strength and weaknesses of measurement instruments and so on, should be obligatory in every empirical paper. Empirical research is always a more or less adequate abstraction of the studied phenomenon and, thus, a lot of issues need to be discussed.

Hopefully, this commentary will also help to stimulate the discussion about these essential issues in quantitative dream research and hopefully result in an overall improvement of the scientific rigor of the studies in our field.

References

Beaulieu-Prevost, D., & Zadra, A. (2007). Absorption, psychological boundaries and attitude towards dreams as correlates of dream recall: two decades of research seen through a meta-analysis. Journal of Sleep Research, 16, 51-59.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale: Lawrence Erlbaum.

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and nonrandomised studies of health care interventions. Journal of Epidemiology and Community Health, 52, 377-384.

Hoffman, C. (2013). Research articles in *Dreaming*: A review of the first 20 years. Dreaming, 23, 216-231.

Kramer, M. (2000). Dreams and psychopathology. In M. H. Kryger, T. Roth & W. C. Dement (Eds.), Principles and practice of sleep medicine (pp. 511-519). Philadelphia: W. B. Saunders.

Kramer, M. (2010). Dream differences in psychiatric patients. In S. R. Pandi-Perumal & M. Kramer (Eds.), Sleep and mental illness (pp. 375-382). Cambridge: Cambridge University Press.

Kramer, M., & Nuhic, Z. (2007). A review of dreaming by psychiatric patients: an update. In S. Pandi, R. Ruoti & M. Kramer (Eds.), Sleep and psychosomatic medicine (pp. 137-155). New York: Taylor and Francis.

Kramer, M., & Roth, T. (1978). Dreams in psychopathologic patient groups. In R. L. Williams & I. Karacan (Eds.), Sleep disorders: diagnosis and treatment (pp. 323-349). New York: John Wiley & Sons.

Punamäki, R.-L. (1998). The role of dreams in protecting psychological well-being in traumatic conditions. International Journal of Behavioral Development, 22, 559-588.

Schredl, M., Paul, F., Reinhard, I., Ebner-Priemer, U. W., Schmahl, C., & Bohus, M. (2012). Sleep and dreaming in patients with borderline personality disorder: A polysomnographic study. Psychiatry Research, 200, 430-436.

Schredl, M., & Reinhard, I. (2008). Gender differences in dream recall: a meta-analysis. Journal of Sleep Research, 17, 125-131.

Schredl, M., & Reinhard, I. (2011). Gender differences in nightmare frequency: a meta-analysis. Sleep Medicine Reviews, 15, 115-121.

Schredl, M., & Reinhard, I. (2012). Frequency of a romantic partner in a dream series. Dreaming, 22(4), 223-229.

Schredl, M., Wittmann, L., Ciric, P., & Götz, S. (2003). Factors of home dream recall: a structural equation model. Journal of Sleep Research, 12, 133-141.

Stumbrys, T., Erlacher, D., Schädlich, M., & Schredl, M. (2012). Induction of lucid dreams: A systematic review of evidence. Consciousness and Cognition, 21(3), 1456-1475.

Yu, C. K.-C. (2013). The structural relations between the superego, instinctual affect, and dreams. Dreaming, 23(2), 145-155.