Organizing Scientific Thinking Using the QuALMRI Framework

Written by Kevin Ochsner and modified by others.
Based on a scheme devised by Steve Kosslyn.

This handout outlines a structured process for generating, asking, evaluating and answering scientific questions. This process, denoted by the acronym QuALMRI, can be used to organize and plan your own research, outline your writing about it, and help understand the research of others. QuALMRI stands for: **Qu**estion, **A**lternative hypotheses, **L**ogic & design, **M**ethod, **R**esults, and **I**nferences. Importantly, QuALMRI can help you make clear what the *question* is that you are asking, and help you to relate it directly to some means of testing it. Furthermore, this process is useful not just in psychology and science, but in any endeavor in which data is used to discriminate between alternative arguments.

This handout has 4 parts: 1) a detailed description of the 5 main steps of the QuALMRI process; 2) a streamlined blank template for use in outlining; 3) a QuALMRI example outline of a published research article; and 4) a copy of the research article outlined in the example.

Part 1: QuALMRI in Depth

I. Begin with a Question: i.e. what do you want to know about? All research should be motivated by a clearly defined question or set of questions that the research seeks to address.

A. Questions can be considered at two levels:

1. Diffuse questions provide a "bigger-picture" motivation for research, such as, "how does gender matter in social interaction?" These questions can not be addressed in single experiments, but rather are answered by considering patterns of data across different studies, each of which addresses a more specific question.

2. Specific questions are "bite size" pieces of larger, more diffuse, ones. For example, given a general interest in gender and social interaction, a more specific, more *focused* question might be, "how do men and women offer different types of verbal and nonverbal feedback during conversation?" This question could be refined further by focusing more closely on examination of specific types of feedback that are offered particular situations.

B. Make clear the connection between diffuse and specific questions. Usually a diffuse "big picture" question is made clear at the beginning of an article, and specific smaller questions to be addressed in individual experiments are also identified. Clarifying the connection between them is essential for mapping out the motivation behind the research. As an experimenter, this entails making clear how your specific questions relate to the more diffuse motivating issues; as a reader/evaluator of other's work, it entails extracting this information and making sure that the specific questions addressed in each experiment meaningfully relate to the "bigger picture."

C. Good questions are motivated by theories, and experiments addressing such questions have implications for the correctness of the theory. Just as we can not explore a large issue in just one study, we can not design a single experiment to test a comprehensive theory all at once. Instead, we have to test the theory bit by bit, using simple experiments that each alone address only small questions, but together have implications for the correctness of the theory as a whole. In addition, interesting questions can be posed such their answers support only one of a number of competing theories.

D. Consider relationship to converging evidence. Remember that a single specific question and an individual experiment don't have to provide an answer to a whole "big picture" question all at once. Although the ultimate goal may be to solve a large puzzle, other research might be available that provides many important pieces, and the single experiment you're designing or evaluating may provide but a single piece. In general, a balanced approach to research involves the use of many different experimental methods and subject populations to help *converge* upon a single explanation, theory, or inference: the results of each study, when considered in isolation, may be subject to different interpretations, but when considered in concert with other supporting evidence, point towards a single conclusion. Remember, however, that it takes time to accumulate this evidence, and that it may be necessary to conduct many experiments before the larger puzzle is complete. Steady, carefully planned, and incremental research wins the race, and along with converging evidence is a particularly powerful research technique, especially when employed in the service of focused research questions.

E. Check to make sure you've got a clear question. Sometimes, however, research is conducted without a clear question in mind; if you are having trouble figuring out exactly what question is being addressed in a paper you read or for a study you're designing, it could be that the authors (which may be you) were similarly confused. In such cases it is difficult to determine what the results of the experiments mean - without knowing what the question is that motivated a given experiment, we have no clear means of inferring what the results of that experiment mean. Always remember that, "what counts as an answer depends on the question you ask." If you're unclear about what is your question, you can't know what to count as an answer.

NOTE: many important scientific advances come unexpectedly and unpredictably, which may at first seem at odds with the idea that steady incremental research is a good way to proceed. Some researchers do take an exploratory approach, and test out "wild" ideas just to see what will happen. However, this does not mean that the experiments conducted to test these speculative ideas were not well planned or carefully designed for a purpose. When unexpected results are produced, one has to be open to considering their potentially novel implications, and know when and how to explore them further.

II. Alternative Hypotheses: The questions addressed in particular experiments should have more than one possible answer; we do experiments to figure out which one(s) are right.

A. There are two broad ways of presenting alternative hypotheses

1. Your/the main hypothesis could be correct or incorrect Sometimes

the alternatives may simply be that something either will or will not happen. For example, one might hypothesize that women are sensitive to negative nonverbal feedback during conversation, whereas men are fairly insensitive to this information. One alternative to this hypothesis is simply that this prediction is incorrect.

2. There are different types of alternative hypotheses, only some of which may be considered within a single experiment. In this case, your hypothesis might be one of a whole class of alternatives: For example, one might hypothesize that negative emotional feedback during conversation can result in reward, punishment, indifference, or other outcomes depending on the manipulation of contextual and other variables that influence that nature of the conversation and the people involved.

B. Order of presenting hypotheses In writing, you should try to spell out all the possible answers to your question and show how they relate to the issue in the following order:

1. F**irst provide the hypothesis that you think is correct,** and explain why.

2. Then present the other alternatives. Be sure to make clear how your experiment will tease apart each hypothetical, alternative outcome. The important thing is that each possible alternative bear directly on the question at hand, and that the question is posed such that its possible answers are theoretically meaningful.

III. Logic & Design: The next step is to consider the basic design of your experiment keeping in mind that the results of your experiments should support one or another alternative hypothesis. The basics of the design include:

A. Specification of dependent and independent variables: Dependent variables are what you measure (e.g. time to react in a conversation, attitudes and their change, etc.) and independent variables are what you manipulate (e.g. the kinds of conversational requests that might elicit emotional reactions). An independent variable might include manipulation of subject groups, such as men vs. women, high vs. low self-monitors, young vs. old or psychiatric patients vs. controls. The use of special subject populations should tell you something interesting about the phenomenon you're measuring. For example, old subjects might respond with less extreme emotional reactions than young subjects, and might be more sensitive to feedback about the appropriateness of their conversational styles. Including young and old subjects in the experimental design allows this hypothesis to be tested.

B. Operational definitions of variables of interest: An operational definition specifies the "operations" in your experiment that you will count as a measure of the "thing" you're interested in. Thus in the conversation example offered above, we might want to devise measures of the intensity and kind of nonverbal emotional feedback given, their impact on other people, and the outcomes that follow from them. The way in which we measure these variables becomes our definition of them within the context of our experiment. In general, you should try to measure and/or control any variable that you think might meaningful influence the outcome of your experiment(s).

C. Deductive logic statements for your question specifying how an experimental outcome will follow from particular alternative answers to your

question: These statements are in the form of, "If (hypothesis X) is true, then (manipulating independent variable x should lead to result y)." One of these statements can be made in support of each alternative hypothesis.

IV. Method After coming up with the basic design you have to specify the nuts-and-bolts of your experiment. This includes:

A. Most importantly, specification of each independent and dependent variable used in your design. This means (depending on the specifics of your experiment) stating exactly

1. who the subjects will be, how they were/will be recruited and

2. what kind of stimuli or questionnaires will be used, how many , and how they relate to each experimental condition defined by your independent variables.

B. Explanation of the procedure that will be used. This includes a description of the sequence of events that occurs when subjects participate in the experiment and should specify:

1. the instructions they are given,

2. what they see, when, for how long, and in what order;

3. how data is collected, including what keys subjects press to indicate which choices, and when,

4. and finally, about how long it the entire **experimental procedure will take.**

V. Results

A. Presentation of results should be in order of importance and relevance to the initial question, and should indicate what values of your variables were found in conducting your experiment. Tables and graphs are often used to clearly present data.

B. Remember, however, that one should **write clear descriptions of the data,** whether it is presented in the text, shown in tables, figures or charts; it is never safe to assume that data speak for themselves. When writing, always keep in mind that one is trying to make sure that the relationship between the results and the initial question is kept clear. The results section is part of the larger story being told by an experiment, and it is important to make sure that the results presented are providing answers to the questions you posed in the introduction. For example, if negative nonverbal feedback more often results in punishment than in reward, then state that in the text, present it in a table or graph (its easier to remember trends and patterns when presented visually as well as verbally) and refer readers to that graph or table in the text.

VI. Inferences After examining the results, it should be possible to use the experimental logic to infer that one or more alternative hypotheses are supported by the data.

A. Consider first the inferences most directly implied by the results and most relevant to the questions at hand, in order of importance. If it is difficult to determine

how the results lead to inferences about the questions, then either the question was poorly defined, the logic was flawed, the method was faulty, or you're tired.

B. When evaluating someone else's work or even one's own, **be careful to discriminate between the inferences that the authors (which might be you) of the study *wish* to draw, and those that are *warranted* by the results.** Try to find flaws in the experimental design, faulty logic, flawed procedures, fuzzy questions, biased sampling of subjects, etc., that limit the ability to draw meaningful inferences from the data.

C. Make suggestions as to how to fix flaws, overcome them, or follow-up on them in subsequent experiments. Note that not all limitations damn a study to the scrap heap; *all* studies are limited in *some* way (remember that each study is addressing only a small piece of a larger puzzle, and so *by definition* is limited in scope) and what's most important is whether and how limitations restrict the inferences that can be drawn about the question that is of interest. If the results are confusing or can't answer the question in a reasonable way, then it may be time to start over at the drawing board.....which is not necessarily BAD - as noted earlier, many important advances come unexpectedly, and in ways that suggest confusing answers to the questions that initially motivated the research. Part of the art of science lies in knowing whether a "weird" result is truly weird and unlikely to replicate, whether it is due to faulty experimental design, whether it means that unwarranted assumptions have been made about the phenomenon being studied, or if it means that a particular line of research should be abandoned, and a new one devised that explicitly follows-up on this new and intriguing finding.

Part 2: QuALMRI Template

I. Question

A. Diffuse, or "big picture" question:

B. The specific question(s) addressed in the research:

C. The connection between the two:

II. Alternative Hypotheses

A. Your/main hypothesis:

B. Other alternatives:

III. Logic & Design

A. Specification of dependent (DV) and independent (IV) variables:

B. Operational definitions of variables of interest:

C. Deductive logic statements for your question specifying how an experimental outcome will follow from particular alternative answers to your question:

IV. Method

A. Realization of each independent and dependent variable:

1. Subjects:

2. Stimuli or questionnaires:

B. Procedure:

1. Instructions:

2. What they see, when, for how long, and in what order:

3. Data Collection:

4. Length of entire experimental procedure:

V. Results

A. Presentation of results in order of importance and relevance to initial question(s):

B. Descriptions of the data shown in tables, charts, etc., as necessary:

VI. Inferences

A. Inferences most directly implied by the results and most relevant to the questions at hand, in order of importance:

B. Discriminate between the inferences that the authors (which might be you) of the study *wish* to draw, and those that you think are *warranted* by the results, by identifying potential flaws and limitations in any stage of the experiment:

C. Suggestions as to how to fix flaws, overcome them, or follow-up on them in subsequent experiments.

Part 3: QuALMRI EXAMPLE

This is a QuALMRI outline of:

Swann, W. B., Hixon, J. G., & De La Ronde, Chris. (1992). Embracing the bitter "truth": negative self concepts and marital commitment. Psychological Science, 3(2), 118-120.

I. Question

A. Diffuse, or "big picture" question: How does the self-concept influence behavior? Why? How do variations in our self-concepts lead us to want different things

in relationships?

B. The specific question(s) addressed in the research: William Swann's self-verification theory predicts that people desire confirmation of their self-concept from their partners. This theory leads him to ask: do people who have negative self-concepts look for confirmation of this view through unfavorable appraisals of themselves by their partner? Similarly, will people who have positive self-concepts seek verification of this view in a positive appraisal by their partner?

C. The connection between the two: Self-verification theory proposes that self-concepts allow us to predict and control how others will react to us in the future, and our self-concepts are derived largely from feedback from others abstracted across time. Consequently, if people aren't giving us feedback which matches how we have come to see ourselves, then our predictions about how they should be reacting to us will be incorrect. Being wrong about our self-concept is aversive, and is avoided; therefore, we'll prefer to interact with, and have relationships with, people whose behavior is predictable and in our eyes, accurate.

II. Alternative Hypotheses

A. Your/main hypothesis: People with negative self concepts either will be in relationships with partners who evaluate them negatively.

Similarly, people with positive self-concepts will enter into relationships with partners whose appraisals confirm their self-concept through positive evaluation.

B. Other alternatives: Both predictions could be incorrect; people with negative self concepts either *will not* be in relationships with partners who evaluate them negatively. Likewise, people with positive self-concepts *will not* enter into relationships with partners whose appraisals confirm their self-concept through positive evaluation.

Furthermore, other theories hold that people generally tend to seek out positive evaluations, which leads to the alternative hypothesis that regardless of whether you have a positive or negative self-concept, you'll seek out positive evaluations from your love partner.

III. Logic

A. Specification of dependent (DV) and independent (IV) variables: The DVs were self-concept ratings for self and partner and ratings of marital commitment for self. There was no IV.

B. Operational definitions of variable of interest: Self-concept defined as the sum of scores on different scales of the Self-Assessment Questionnaire (SAQ), which includes questions assessing intellectual, athletic, social, artistic, and musical ability as well as physical attractiveness. Commitment was defined as the sum of scores on scales rating, "intentions, feelings and actions regarding their relationship."

C. Deductive logic statements for your question specifying how an experimental outcome will follow from particular alternative answers to your

question: *If* self-verification theory is correct and people seek partners who verify their self concept, *then* people who rate themselves unfavorably as measured by the SAQ should be more committed to spouses whose SAQ ratings for them are similarly unfavorable. The converse should hold for those with positive self-concepts. Alternatively, if people seek only positive feedback from their partners, then regardless of self-concept, subjects should be more committed to partners offering positive evaluations of them.

IV. Method

A. Realization of each independent and dependent variable:

1. Subjects: 95 married couples who went to horse ranches and shopping malls, mean age 32 yrs, have been together for 9 yrs on average, and married for 6 of them.

2. Stimuli or questionnaires: As discussed above, the SAQ and other commitment scales were administered.

B. Procedure:

1. Instructions: Subjects were told that the experiment concerned "the relation between personality and close relationships" and completed the questionnaires while seated at opposite ends of a table.

2. What they see, **when, for how long, and in what order:** the SAQ and other scales were completed at subjects own pace.

3. Data Collection: written response on questionnaires were collated & analyzed.

4. Length of entire experimental procedure: not given.

V. Results

A. Presentation of results in order of importance and relevance to initial question(s): Results supported the main hypothesis: subjects were more committed to partners who offered similar appraisals of their self-concepts.

B. Descriptions of the data shown in tables, charts, etc., as necessary: This results is illustrated in Table 1, p. 119: commitment were greater when the appraisals of both the subjects and their partners were both negative ($\underline{M}$ = 52.4) or positive ($\underline{M}$ = 58.7, than when their appraisals differed ($\underline{M}$ = 43.8 for self negative, partner favorable; $\underline{M}$ = 52.0 for self positive, partner unfavorable).

VI. Inferences

A. Inferences most directly implied by the results and most relevant to the questions at hand, in order of importance: Self-verification theory is supported; people do seem to prefer feedback that verifies or confirms their self-concepts.

Additional data analysis suggested that verification was accompanied by a feeling that their partners, "helped them understand themselves," did not differ for men and women, and in the case of desiring unfavorable feedback was not attributable to a desire to win over a recalcitrant lover, belief that the partner was especially perceptive, or belief that partners would help them improve themselves. If you think little of yourself, you want people to tell you this; positive feedback just could not be believed.

B. Discriminate between the inferences that the authors (which might be you) of the study *wish* to draw, and those that you think are *warranted* by the results, by identifying potential flaws and limitations in any stage of the experiment: A few caveats to these conclusions might be offered, however. First, the subjects used in this experiment might not be representative of the entire population - there might be something about couples who go to ranches together or who shop together in middle-class suburban Texas which makes them more likely to fit Swann's predictions. Would the urban poor, or people in other cultures feel the same way? Second, the correlational design can not tell us whether people had negative (or positive) self-concepts first, and then sought out someone who gave them congruent feedback, or whether their negative (or positive) self-concept is in some way a result of negative (or positive) feedback. The causal, developmental chain in this process should be a topic of future study. It would also be interesting to consider the applicability of the present results and methodology to the understanding of commitment to abusive relationships: does one stay because the abused partner believes that the abuse is deserved?

C. Suggestions as to how to fix flaws, overcome them, or follow-up on them in subsequent experiments. To test the generality of the present findings, the experiment could be replicated in other populations located in other contexts. For example, couples that don't tend to shop or travel much together could be recruited and screened through newspaper ads. To address the question of causality, a longitudinal study could assess self-concept in couples that have just met, and see how their self and partner ratings change over time as a function of the success or failure of the relationship. The subjects in the present experiment could also be followed-up later to see if their ratings are stable, and to examine the consequences of self-verifying of self-non verifying feedback in the long-term.