Argument Quality and Persuasive Effects: A Review of Current Approaches

Daniel J. O'Keefe, University of Illinois at Urbana-Champaign
Sally Jackson, University of Arizona

This essay reviews current approaches to the study of the persuasive effects of variations in argument quality. The essay initially sketches the importance of considering the role that argument quality plays in persuasion. It then describes three general approaches to the study of argument quality, arguing that each has important shortcomings. These defects are seen to have a common underlying basis, namely, the lack of an independently-motivated normative account of argument quality. It is concluded that progress in understanding the role of argument quality in persuasion will require such an account.

The Importance of Argument Quality

Persuasive messages commonly contain (implicit or explicit) arguments in favor of the advocated position. These arguments may vary in number, in content, in how they are ordered in the message, and so forth. A number of these dimensions of argument variation have received empirical attention as possible influences on the success of persuasive messages (see, e.g., Bridges & Reinard, 1974; Chaiken, 1980; Glikinson, Paulson, & Slikkink, 1954). One particularly notable way in which message arguments may vary is quality. That is, a given argument might be a normatively good argument (a high-quality argument) or might be a normatively poor argument (a low-quality argument).

There are at least two reasons for considering more closely the role that argument quality variations may play in persuasion. One is simply the manifest value of understanding that role. Consider: Under what circumstances will genuinely better arguments be likely to prevail? What factors influence the effect that argument quality variations have on persuasive outcomes? To which aspects of argument quality are message receivers likely to be especially attentive? Questions such as these are certainly worth pursuing, and are at least as deserving of attention as questions focused on what are arguably rather more superficial aspects of persuasive messages.

A second reason is the important role that the concept of argument quality plays in recent theoretical approaches to persuasion. In particular, argument quality figures significantly in the elaboration likelihood model of persuasion (Petty & Cacioppo, 1986).

The elaboration likelihood model (ELM) proposes a broad distinction between two different general processes of persuasion. Which persuasion process is activated in any given circumstance is seen to depend upon the degree of issue-relevant thinking ("elaboration") engaged in by the receiver. The degree of elaboration is taken to be a function of such factors as the personal relevance of the topic to the receiver (with greater relevance, increased elaboration is likely), the presence of distraction in the persuasion setting (with increasing distraction, there is likely to be reduced elaboration), and so forth.

According to the ELM, when the receiver does not engage in extensive issue-relevant thinking, "peripheral routes to persuasion" are said to be engaged. In such a circumstance, the outcome of persuasive efforts is taken to commonly depend upon the receiver's use of simplifying decision principles ("heuristics"); for example, receivers might be guided by the degree to which they like the communicator, or by their perception of the communicator's credibility. By contrast, when the receiver does engage in extensive issue-relevant thinking, a different persuasion process is engaged, the "central route" to persuasion. In such a circumstance, the outcome of a persuasive effort is seen to depend on the results of the receiver's close scrutiny of the message's contents—and hence to depend upon (inter alia) the quality of the arguments contained in the message. Thus the general ELM image is that as issue-relevant thinking increases, the impact of peripheral considerations (such as the receiver's liking for the communicator) on persuasive outcomes will decrease, and the impact of argument quality (the degree to which the message's arguments are cogent or specious) on persuasive outcomes will increase (see Petty & Cacioppo, 1986, pp. 141-172). Argument quality is thus a key variable in the ELM's depiction of how persuasion works.

Despite the apparent importance of considering the effects of argument quality variations on persuasive effects, the research literature on persuasion has seen relatively little concerted attention given to the conceptualization of argument quality (even within the ELM). That is, there is no general abstract characterization of what constitutes argument quality, no clear identification of different facets of argument quality, no thorough systematic analysis of argument quality.

The absence of a general conceptual treatment of argument quality has been no barrier to empirical studies of argument quality, however. Of course, in order to study empirically the persuasive effects of argument quality variations, researchers must have some means of establishing argument quality variations. The research literature contains three different approaches to the operationalization of argument quality. In the next section, we critically discuss these approaches, arguing that each has important defects.

Current Approaches to Argument Quality

In operationalizing argument quality variations, researchers have commonly followed one of three procedures: (1) pretesting messages for persuasive effectiveness under conditions of high message scrutiny; (2) obtaining participant ratings of argument quality; and (3) creating unsystematic message variations that might be taken to reflect argument quality variations. Upon close examination, each of these approaches will be seen to be unsatisfactory. Although the particular defects will differ from one approach to another, a common underlying
problem will emerge: the lack of an independently-motivated normative account of argument quality.

**Pretested Effectiveness**

One way in which argument quality variations have been empirically created is by pretesting messages for their persuasive effectiveness under conditions of high message scrutiny. This procedure is especially associated with ELM research. Argument quality plays a prominent role in the ELM, so perhaps it is not surprising that the ELM has a correspondingly explicit treatment of the operationalization of argument quality.

**The pretested-effectiveness procedure.** The recommended ELM procedure involves initially having potential arguments rated for persuasiveness by pretest participants. Then messages composed of high- and low-rated arguments are presented to other pretest participants, who report their thoughts while receiving the messages. Messages that elicit predominantly favorable reactions are deemed to contain high-quality arguments; those that elicit predominantly unfavorable reactions are taken to contain low-quality arguments (see Petty and Cacioppo, 1986, p. 32; see also p. 54n4). This procedure has been followed in a number of ELM-related studies of persuasive effects (e.g., Neimeyer, MacNair, Metzger, & Courchaine, 1991; Petty, Cacioppo, & Goldman, 1981; Petty, Cacioppo, & Heesacker, 1981).

A key defect in the pretested-effectiveness procedure. This procedure has one central defect: it assumes, rather than shows, that argument quality is responsible for persuasiveness under conditions of high elaboration. That is, rather than providing evidence about the role of argument quality, this procedure begging the question of the role of argument quality in persuasion (see O'Keefe, 1990, pp. 110-111).

As a way of displaying this problem, consider the question "what is it about the strong-argument messages that makes them persuasive under conditions of high elaboration?" One is naturally tempted to answer "it's the strong (high-quality) arguments they contain—that's what makes them persuasive." But this answer is mistaken. The messages were constructed (through pretesting) precisely so that, under conditions of high elaboration, favorable reactions would be obtained. That is, argument quality is operationally defined by argument effects under conditions of message scrutiny. Given this way of defining high- and low-quality-argument messages, it is guaranteed that the "strong-argument" messages will be persuasive under conditions of high elaboration. If an experimental message did not already produce favorable reactions (in pretesting), it would not be labeled a "high-quality" message.

In short, rather than showing that argument quality is responsible for persuasiveness when elaboration is high, the ELM procedure assumes (in a definitional way) such a role for argument quality. Thus the ELM begs the question of argument quality's role in persuasion, by virtue of the way in which argument quality is operationally defined.

**The danger of the "argument quality" label.** Petty and Cacioppo do explicitly recognize that they "have ignored the specific qualities that render some arguments cogent and others specious" (Petty & Cacioppo, 1986, p. 32). But they believe that they have simply "postponed the question of what specific qualities make arguments persuasive by defining argument quality in an empirical manner" (p. 32).

That is to say, in their view, the "argument quality" label is simply a benign placeholder.

But in fact the ELM's use of the "argument quality" label is dangerous and misguided, because it invites bad thinking. In particular, it encourages one to suppose that we understand what it is about those messages that lead to the observed effects—namely, that the messages vary in "argument quality." But, as just seen, this is incorrect.

It will be useful to erase the unhappy consequences of having had these message manipulations already labeled as "argument quality" manipulations. So as not to beg the key questions, the "strong-argument" and "weak-argument" messages (used in ELM research) can be called "S messages" and "W messages," respectively. Some things are already known about the effects of S and W messages. For example, generally speaking, S messages are more persuasive than W messages. And the size of this difference in effectiveness varies; as message scrutiny increases, there's a greater difference in persuasive effectiveness between S and W messages than there is when scrutiny is low.

The question that arises, of course, is: what is the characteristic that varies between S and W messages that explains these effects? (As an aside: our discussion here proceeds on the simplifying assumption that there is just one characteristic that varies between S and W messages that explains these effects. This assumption is not required, and indeed is arguably not accurate. But the end of clear discussion will be better served by proceeding as if there is only one distinguishing causally-relevant feature.)

Expressed somewhat differently, the question is: what is the active ingredient differentiating S and W messages?

When the matter is approached in this way, the ELM might be seen to be offering a hypothesis about what this message characteristic is, namely, "argument quality." But the ELM offers no evidence to support this hypothesis. (Certainly the observed effects cannot be evidence for their own explanation.)

In fact, in order to gather evidence bearing on the ELM hypothesis, one will need to have some independent account of "argument quality." To show that argument quality is the active ingredient differentiating (and explaining the differential effects of) S and W messages will require some way of assessing the argument quality of S and W messages, and then showing the appropriate covariation between argument quality and the observed effects. But assessing the argument quality of messages (as distinct from assessing the effects of those messages) obviously requires some independent standard for argument quality.

**Two research questions.** This clarification of the status of ELM claims about argument quality makes it apparent that in fact there are two distinguishable (but easily confused) research questions here. One is the question identified above: what is the characteristic (the active ingredient) that varies between S and W messages that explains the observed effects of the messages? The other research question is: what role does argument quality play in persuasion? (That is, what influence do variations in argument quality have on persuasive outcomes?) It is easy to confuse these two questions, precisely because the ELM has already invoked the "argument quality" label as an answer to the first question. But, as noted, the ELM offers no independent justification for the use of this label; the invocation of the "argument quality" label amounts to a hypothesis—a hypothesis that still awaits evidence.
The ELM does not (and indeed cannot) address either of these two research questions, because of its approach to operationalizing argument quality. The ELM does not address the question of what the active ingredient is that differentiates S and W messages (but instead assumes that "argument quality" is the ingredient); and the ELM does not address the question of what role argument quality plays in persuasion (because its procedure begs the question of what constitutes argument quality).

**Summary.** In short, then, operationalizing argument quality by assessing argument effectiveness under conditions of high message scrutiny is unsatisfactory. Such a procedure can only assume, rather than show, that argument quality is responsible for persuasiveness under conditions of high elaboration, for the procedure has no independent conception of argument quality.

**Argument Quality Ratings**

A second approach to the operationalization of argument quality variations is to rely on participant ratings of the quality of arguments. That is, participants are asked to rate arguments (or collections of arguments, as in a message that contains several arguments) for quality. These quality ratings are then taken to provide a basis for operationalizing argument quality: arguments that participants rate as high in quality are taken to be high-quality arguments, and arguments rated as low in quality are taken to be low-quality arguments.

**The quality-rating procedure.** Quality ratings can be obtained in various ways: participants might be asked to rate arguments individually, or to rate messages (that is, groups of arguments); the wording of the rating scale may vary; and so forth. But the underlying idea is that participants will provide assessments of argument quality. For example, Axsom, Yates, and Chaiken (1987) established argument quality variations by having pretest participants rate a pool of arguments for strength. "The high- and low-quality versions [of the experimental messages] contained the six highest and six lowest rated arguments, respectively, from a pilot study conducted to identify strong and weak arguments" (p. 32). Notably, Axsom et al. (1987) also reported a "manipulation check" in which main-study participants were asked to rate on a 15-point scale "the overall quality of the arguments used by the debater to support his position" (p. 33). Similar invocation of participant ratings of argument quality can be seen in a number of studies, including Andrews and Shimp (1990), Burnkrant and Howard (1984), DeBono (1992), Mackie, Worth, and Asuncion (1990), and Munch and Swasy (1988).

**Defects in the quality-rating procedure.** Procedures such as these have two important weaknesses. First, using participant perceptions of argument quality as a basis for establishing argument quality variations rests on the supposition that receivers' perceptions of argument quality accurately reflect argument quality. To assert, for instance, that participants' ratings provide a "check" on the manipulation is to assume that participants' ratings are relevant to the success of the manipulation, and that if participants did not perceive argument-quality differences then there must not have been argument-quality differences.

This is surely a dubious general assumption. It requires believing that (for example) participants are never fooled by fallacious reasoning; they always know genuinely meritorious arguments when they see them; they are invariably capable of spotting flaws in reasoning, weaknesses in supporting evidence, inconsistencies in arguments, and so forth. If nothing else, a glance at the existing literature on human reasoning and decision-making processes might give one some pause in accepting such premises (see, e.g., Arkes & Hammond, 1986; Garmham & Oakhill, 1994; Nisbett & Ross, 1980). As Johnson-Laird and Shafir (1994) have remarked, "evidence continues to accumulate documenting the violation of even the most essential [normative] principles" of rational inference and decision; indeed, they conclude that "the major psychological discovery about both reasoning and decision making is that normative theory and psychological facts pass each other by. People are not intuitive logicians, intuitive statisticians, or intuitive rational decision theorists" (p. 6).

Second, the supposition that receivers' perceptions of argument quality accurately reflect argument quality is not evidenced. That is, the applications of this procedure have simply taken it for granted that receivers' perceptions of argument quality accurately reflect argument quality. But (especially given that some skepticism about this assumption is plausible) one might justifiably ask for some supporting evidence. Providing such supporting evidence, of course, will require some independent standard of argument quality (some standard apart from participant ratings)--and this is lacking from the quality-rating procedure.

**Summary.** In sum, operationalizing argument quality on the basis of perceived argument quality is unsatisfactory. The procedure rests on a questionable assumption that participants' perceptions of argument quality accurately reflect argument quality; and supporting that assumption would require an independent account of argument quality, which the procedure lacks.

It may be noticed that the two research questions distinguished previously (in the discussion of the ELM's demonstrated-effectiveness procedure) also arise distinctly in the context of this argument-quality rating procedure. First, there is the question of what actually distinguishes those arguments rated as "high quality" and those rated as "low quality." Second, there is the question of what role argument-quality variations play in persuasion.

Using argument quality ratings to operationalize argument quality variations leaves both these questions unexplored. Using argument-quality ratings as the basis for argument-quality manipulations leaves unaddressed the question of what the active ingredient is that differentiates "rated-as-high-quality-argument" and "rated-as-low-quality-argument" messages. And using argument-quality ratings makes it impossible to address the question of what role argument quality plays in persuasion (unless one makes the dubious and unevidenced equation of perceived argument quality with argument quality).

**Unsystematic Message Variations**

A third approach to the operationalization of argument quality has been to manipulate, in a relatively unsystematic fashion, various message features that might be taken to be related to argument quality.

The unsystematic-variations procedure. This way of operationalizing argument quality variations does not proceed on the basis of some well-articulated general conception of argument quality. Instead, the features of arguments (or messages) are varied in ways that the
researcher believes (on some unspecified basis) reflect argument quality.

For example, Hunt, Smith, and Kernan (1985) varied "argument quality" by varying the specificity or tangibility of supporting arguments for a particular brand of computer. "Argument strength was manipulated by exposing subjects to different information concerning the computer's ease of use. Some subjects read a description that presented specific, or tangible, arguments on behalf of the computer's easy handling characteristics. Others were exposed to less specific (more intangible) argumentation . . . The strong-argument (tangible) condition described the computer's easy-to-use function keys that would draw graphs or construct tables, while the weak-argument (intangible) condition suggested merely that one could master the computer's operation in a short time" (p. 452).

A different manipulation was used in Jepson and Chaiken's (1990) study of various messages concerning cancer checkups. They "created strong and weak versions of each cancer-related message. This was accomplished by varying the strength of the figures given in the statistics supporting the arguments" and by having the weak version contain logical errors that "were corrected in the strong version" (p. 74n6).

Still another procedure was employed by Bohner, Chaiken, and Hunyadi (1994), who created a manipulation described as involving "a message containing unambiguous strong, unambiguous weak, or ambiguous arguments" (p. 207). The strong/weak manipulation was based on variations in the number and importance of attributes ascribed to the attitude object. The "unambiguous strong" message described the object as superior on four important attributes and inferior on two unimportant ones; the "unambiguous weak" message described it as superior on four unimportant attributes and inferior on two important ones (see p. 212; also see Chaiken & Maheswaran, 1994, from which these messages were adapted).

**Defacts in the unsystematic-variations procedure.** These various operationalizations of argument quality are plainly unsystematic; they are not based in some larger analysis of argument quality. This makes for two central weaknesses.

First, legitimate questions can arise about the degree to which a given manipulation in fact realizes argument quality variations. Without a general analysis of argument quality, there is no non-intuitive basis for supposing that a given manipulation in fact operationalizes argument quality. Notice, for example, that in order to justify Hunt et al.'s (1985) manipulation as yielding genuine argument quality variations, one will need to display some connection between the particular manipulation and some general characterization of argument quality.

A second, more significant shortcoming is the lack of any larger framework within which to understand such diverse manipulations. Consider, for instance: How are "logical errors" and "attribute importance" related (as means of operationalizing argument quality)? Are these procedures fundamentally the same, or importantly different? Just which facets of argument quality are being manipulated in these various procedures? Should similar or different effects be expected from these different manipulations? Are all facets of argument quality likely to have identical persuasive effects when varied? Plainly, addressing questions such as these will require some systematic analysis of the nature of argument quality.

**Summary.** Operationalizing argument quality variations through unsystematic manipulation of message features, unlike the use of demonstrated argument effectiveness or the use of participant quality ratings, does offer the prospect of illuminating the relationship between specific message variations and persuasive effects. Unfortunately, because it is not based on some general conception of argument quality, the procedure cannot be assured to actually produce argument quality variations, and is unlikely to yield systematic insight into the role of argument quality in persuasion.

**The Underlying Problem.**

The three approaches sketched here are not entirely independent. For example, it is possible to supplement the demonstrated-effectiveness (ELM) procedure by adding elicitation of main-study participant ratings of message quality (see, e.g., Burnkrant & Howard, 1984). And of course both the demonstrated-effectiveness procedure and the participant-rating procedure utilize arguments whose features have been varied in ways that the investigator hopes will eventually result in appropriately-varying arguments (i.e., arguments that differ in demonstrated effectiveness or in rated quality). That is to say, unsystematic message variations lie at the base of all these approaches to operationalizing argument.

And that, in a way, is precisely the root of the problem. All these procedures for operationalizing argument quality suffer from a common underlying flaw: the lack of an independently-motivated normative account of argument quality. The demonstrated-effectiveness (ELM) procedure begs the key questions about argument quality, because it has no effects-independent conceptualization of argument quality; the quality-rating procedure is unable to provide evidence that perceived argument quality reflects actual argument quality, because it lacks an independent standard for argument quality; and the unsystematic manipulation of message features yields only haphazard and disconnected findings, because it lacks any larger conceptual framework within which to understand particular variations.

**Conclusion.**

It should now be plain that any progress in understanding the persuasive effects of argument quality variations will require the articulation of a larger conceptual framework for the analysis of argument quality, that is, some independently-motivated account of argument quality.

Perhaps it will come as no surprise to the reader that we believe it is possible to provide such an account, and more specifically to provide an account of argument quality that can illuminate and guide empirical research on argument quality effects. Although space does not permit a full presentation here, we can briefly say that we believe such an approach is most usefully constructed not on traditional philosophical normative standards for argument (truth of premises, validity of form) but on standards developed from dialectical and pragmatic approaches (e.g., Barth & Krabbe, 1982; van Eemeren, Grootendorst, Jackson, & Jacobs, 1993; Freeman, 1991) and related work in informal logic (e.g., Johnson & Blair, 1983; Walton, 1989; Woods & Walton, 1982). Such standards would not be focussed on "message features" as commonly conceived in the persuasion effects literature, but on underlying message production principles that reflect the procedural obligations associated with good argumentation.

Whether the particular approach we plan to pursue will bear fruit remains to be seen. In any case, however, if progress is to be made in understanding the role of argument quality variations in persuasion, some
independent normative analysis of argument quality will be needed.

Notes

1 A roughly similar image of persuasion has been offered by the heuristic-systematic model (HSM; Chaiken, 1987; Chaiken, Liberman, & Eagly, 1989). Although there are some important differences between these approaches, the ELM and HSM depict a broadly similar role for argument quality in persuasion. However, argument quality plays a larger and more explicit role in ELM theorizing and research—thus the present emphasis on the ELM.

2 This is a variant of what Abeelson (1995, p. 136n5) calls the “knighthood fallacy”: “When we title a treatment factor, there is always a danger that we will proceed to use that label in our thinking and ignore the idiosyncratic nature of the actual treatment manipulation.”

References


Argumentation and Values

Proceedings of the Ninth SCA/AFA
Conference on Argumentation

Editor

Sally Jackson

Associate Editors

Rebecca S. Bjork
G. Thomas Goodnight
Scott Jacobs
James F. Klumpp
Malcolm O. Sillars
Charles Arthur Willard

Sponsored by
The Speech Communication Association
The American Forensic Association
The University of Utah

Publisher of Record:
Speech Communication Association
5105 Backlick Road, Suite E
Annandale, VA 22003

1995