Convergent causal arguments in conversation

Dale Hample
University of Maryland

Katarzyna Budzynska

Follow this and additional works at: https://scholar.uwindsor.ca/ossaarchive

Part of the Philosophy Commons

https://scholar.uwindsor.ca/ossaarchive/OSSA9/papersandcommentaries/28

This Paper is brought to you for free and open access by the Conferences and Conference Proceedings at Scholarship at UWindsor. It has been accepted for inclusion in OSSA Conference Archive by an authorized conference organizer of Scholarship at UWindsor. For more information, please contact scholarship@uwindsor.ca.
Convergent causal arguments in conversation

DALE HAMPLE

Department of Communication
University of Maryland
College Park MD
USA
dhample@umd.edu

ABSTRACT: In theory, flawed arguments are not individually sufficient to justify a conclusion, but several may converge to do so. This is an empirical study of how arguers respond to a series of imperfect causal arguments during a serious conversation. People became less critical of the flawed arguments as more of the arguments appeared. The study gives empirical evidence that ordinary arguers permit sufficiency to accumulate during an extended discussion.

KEYWORDS: Causal argument, conductive argument, convergent argument, conversational argument, face-to-face argument, sufficiency.

1. INTRODUCTION

This investigation is concerned with arguments whose conclusion is supported (or at least pointed to) by several lines of argument. Our community vocabulary for this situation is somewhat variable. An important distinction is whether the individual lines of argument each bear directly on the single overall conclusion, or whether the lines of argument point to one another in some sort of sequence, so that only the last bit of argumentation is supposed to produce the overall conclusion directly. I am concerned entirely with the first case, in which several conceptually independent reason sets each are pointed to the conclusion.

2. CONVERGENT ARGUING

Based on Henkemans’ (2000) discussion, I call this phenomenon convergent argumentation. Since it goes by other terms (e.g., van Eemeren, Grootendorst, Jackson, & Jacobs, 1993, call it coordinately or subordinately compound argumentation) and is subsumed under (or contrasted with) other terms such as complex argumentation and conductive argument, I offer the following diagram for clarity (Fig. 1). The diagram illustrates a central conclusion that is supported by three different arguments. Those arguments have no special relationship to one another, so that the strength of one has nothing to do with the strength of another. Each could be offered as the sole support for the main conclusion. The arguer’s strategy, of course, is that somehow the main conclusion will be more strongly implicated with three arguments than with one.
As others have explored, this situation complicates our analysis of “the” argument. If all three converging arguments are poor (i.e., have serious problems either with the illative issues of acceptability, relevance, and sufficiency, or with the dialectical tier; Johnson, 2000) should we conclude that the main conclusion is poorly supported, or might we consider the possibility that individually flawed reason sets might combine in some way to give more than poor support for the conclusion? Notice that this question is only interesting if all the supporting reason sets are flawed. If one were a perfect deduction with acceptable premises and no dialectical vulnerability, the other reason sets would be irrelevant to the force of the conclusion. Is it possible, though, that several very refutable reason sets might somehow combine to make a conclusion acceptable?

I will leave to others the normative and theoretical questions I have just mentioned. Here, my interest is in how ordinary arguers act in the face of multiple flawed reason sets bearing on one conclusion. Argumentation on complicated topics often contains a number of attempts to justify a standpoint. People will try one bit of proof, then another, and this will play out naturally in the course of a serious conversation. Even when the individual lines of proof have observed and fundamental flaws, it may be that as the conversation progresses, participants begin to grant increasing plausibility to the general conclusion, and cease to reject new (weak) supports with much energy; this is the central hypothesis of the study.

3. METHOD

72 undergraduates at the author’s institution participated in argumentative conversations on the topic of gun control. In every case, the other conversant was a confederate who had been trained to insert flawed causal arguments into the discussion. 11 confederates participated in the study. These circumstances were unknown to the participants until the debriefing. The conversations were videotaped and lasted for about 5 minutes.

68% of respondents were women. The sample’s mean age was 20.4 years ($SD = 3.2$). 58% self-reported Euro-American ethnicity, 19% African-American, 7% Asian-American, and the remainder were scattered among various nationalities or combinations of ethnicity. The sample’s mean self-reported high school grade point average was 3.6 on a four point scale ($SD = .33$) and their college grade point average was 3.3 ($SD = .38$).
Respondents participated in order to earn minor extra credit in communication classes. The study was approved by the author’s Institutional Review Board.

Both conversants were directed to role play the part of a U.S. Senator. The pretense for the conversation was that the two “Senators” had come together privately to work out some possible national legislation regulating handgun ownership in the U.S. Both were provided briefings by their fictional staffs. The briefing materials consisted of about 4 pages (single spaced) containing actual quotations taken from government documents, primarily testimony given at a 2005 U.S. House of Representatives committee hearing on gun control. The briefing materials indicated that the two Senators had opposing views, and all of the actual participants were assigned to the standpoint that ordinary citizens should not be allowed to own handguns. The conversations took place at a conference table in our lab, and each Senator sat behind a nameplate (either Senator Jones or Senator Smith). Confederates were instructed to address the participant as “Senator” periodically. These steps were all taken in order to promote the seriousness of the conversation. The public business scenario, the semi-formality of the conversation, and perhaps most importantly, the provision of actual information to be used in the discussion, together resulted in far more substantial arguments than when participants are just invited to argue about some topic on which they may be lightly informed.

Participants were given as much time as they wished to study the briefing materials. Many participants underlined or highlighted various portions of the briefing. They also responded to some demographic questions, and then went into the main part of the lab for the conversation, where they met the confederate.

The confederates’ briefing largely duplicated that of the participants, with the important exception that it concluded with a list of 10 weak causal arguments. These were paraphrases of testimony at the same U.S. House committee hearing. Confederates were instructed to insert as many of these arguments as possible into the conversation. Altogether, 306 instances of these arguments appeared in the 72 conversations.

The arguments are included here in the Appendix. They were constructed in such a way as to fail at least one of the critical questions tests in Walton, Reed, and Macagno’s (2008) analyses of cause to effect, abduction, or correlation to cause argument schemas. Among other flaws, these arguments mistake correlation for causality, try to justify a national policy on the basis of a single city’s experience, or project another country’s results onto the U.S. Readers can see for themselves that on a number of grounds, no one of the arguments in the Appendix is a decisive reason establishing the confederate’s standpoint, namely that handguns should be freely available to U.S. citizens. All offer some sort of causal analysis of the relationship between handgun ownership/restriction on crime. Confederates were trained not to press these arguments once the participant had declined to pursue them because we wanted the participants to be in control of when the line of argument terminated, how it was developed, or even whether it was explored at all.

The portions of the videotapes containing the confederates’ flawed arguments were identified and transcribed. Each transcribed segment began with the confederate’s presentation of the argument, and continued until that line of argument was complete.

The transcribed segments were then coded by the 11 confederates. The coding system was taken from Hample, Jones, and Averbeck (2009). The participant’s behaviors throughout each segment were classified into one of five categories: (1) accepted the ar-
argument; (2) ignored the argument; (3) rejected the argument without giving a reason; (4) rejected the argument with explicit refutation; and (5) took up the argument and revised it to make it sounder. The highest appropriate code for a given segment was recorded.

The codes are in order of sophistication. Given that the arguments were all known to be flawed, the least sophisticated response was to accept it. Ignoring it—most commonly by changing the topic—gives no real information about what the participant thought, but is intermediate between acceptance and refutation. Rejecting the argument with an objection or counter-argument shows more argumentative ability than just rejecting it out of hand. And finally, while rare here, revising the confederate’s argument to make it stronger (perhaps prior to refutation) is most advanced of all.

Reliability of the coding was assessed by regarding the system as interval and each coder as an independent indicator of the true score. Thus the coders became “items” in the usual Cronbach’s alpha analysis of a multi-item instrument. Cronbach’s alpha for the 11 coders was .88. This is well above the usual minimum requirement of alpha = .70.

4. RESULTS

The overall mean code was 3.48 (SD = .74). The average response was thus midway between 3 (rejects without reason) and 4 (rejects with explicit refutation). In other words, respondents generally recognized that the confederates’ arguments were, in fact, flawed. Neither high school GPA (r = -.07, p = .57) nor college GPA (r = .08, p = .54) predicted the mean response. However, age did have a significant effect: r = -.33, p < .01), indicating that older students were more accepting of the weak arguments. Men (Mean = 3.63) and women (Mean = 3.51) did not differ significantly in the sophistication with which they responded to the flawed causal arguments (t = .71, df = 70, p = .48).

<table>
<thead>
<tr>
<th>Position</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>72</td>
<td>3.55</td>
<td>.67</td>
</tr>
<tr>
<td>2</td>
<td>70</td>
<td>3.57</td>
<td>.62</td>
</tr>
<tr>
<td>3</td>
<td>66</td>
<td>3.51</td>
<td>.77</td>
</tr>
<tr>
<td>4</td>
<td>53</td>
<td>3.50</td>
<td>.78</td>
</tr>
<tr>
<td>5</td>
<td>33</td>
<td>3.08</td>
<td>.91</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>3.44</td>
<td>.61</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>3.45</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1. Mean Responses to Flawed Arguments By Conversational Position

The main question of interest is whether participants became more accepting of the arguments as the conversations progressed. We kept track of which confederate argument
appeared first, second, and so forth. The largest number of these arguments in any conversation was 7. Table 1 shows the mean responses for each conversational position. As Table 1 shows, the overall mean response does not change very much through the first few arguments.

Critically for the prospect of testing this result, however, the number of conversations having higher numbers of arguments declines as we move to higher numbers of arguments. To test whether or not mean responses differ, we must analyze the very same respondents. If we wanted to compare the responses to arguments 1, 2, and 3, therefore, we would need to restrict ourselves to the 66 respondents who experienced three or more such arguments.

The best balance between sample size and number of flawed arguments appeared to be with five arguments \((N = 33, 165 \text{ arguments})\). The means for these respondents are in Table 2. Repeated measures analysis of variance was conducted. The sphericity assumption was violated in the data set \((\text{Mauchley}'s \ W = .245, \chi^2(9, N = 33) = 42.82, p < .001)\), so the Greenhouse-Geisser correction was applied. This resulted in \(F = 3.295, df = 2.54, p < .05, \text{ partial } \eta^2 = .09\). Table 2 shows that as we proceed through a sequence of weak arguments, the general response becomes lower (i.e., more accepting), apart from a bump at argument position 4. Both a linear \((F = 6.43, df = 1, p < .05)\) and a cubic effect \((F = 18.67, df = 1, p < .001)\) were significant.

<table>
<thead>
<tr>
<th>Position</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>33</td>
<td>3.60</td>
<td>.56</td>
</tr>
<tr>
<td>2</td>
<td>33</td>
<td>3.46</td>
<td>.68</td>
</tr>
<tr>
<td>3</td>
<td>33</td>
<td>3.31</td>
<td>.87</td>
</tr>
<tr>
<td>4</td>
<td>33</td>
<td>3.58</td>
<td>.62</td>
</tr>
<tr>
<td>5</td>
<td>33</td>
<td>3.08</td>
<td>.91</td>
</tr>
</tbody>
</table>

Table 2. Mean Responses to Flawed Arguments by Conversational Position for Respondents Exposed to 5 or More Flawed Arguments

A test for only four argument positions fails to show the same effect, as inspection of Tables 1 and 2 might suggest. For this data, the sphericity assumption is not violated, and \(F = .18, df = 3, p = .91\). Sample size for this test was 53, and the means for the four argument positions were respectively 3.58, 3.53, 3.51, and 3.50.

5. DISCUSSION

This investigation shows that as conversations moved through a sequence of five flawed causal arguments, people became less critical of them. This was entirely sensible behavior. Readers should notice that all of the arguments in the Appendix have some probative merit. In isolation, none is sufficient to justify the confederate’s standpoint. But unlike
extreme fallacies—e.g., an invalid categorical syllogism—these arguments are merely weak, not completely valueless. The arguments in the Appendix are all relevant to the main conclusion and all ought to have been judged as being formed from acceptable premises. The essential problem for the arguments considered individually was sufficiency. In combination, they began to approach that criterion, at least insofar as our participants’ responsive behaviors can inform us.

Those responses changed mainly on the occasion of the fifth argument here, but there is little reason to suppose that the fifth position has any generalizable importance. Assuming that people’s behaviors are approximating rationality in some degree, the point in the conversation where resistance begins to soften ought to be influenced by the quality of the flawed arguments. If the arguments have barely any probative value at all, the transition should be later. If the arguments have relatively minor flaws, arguers might feel at an earlier point that sufficiency has been approached or achieved.

It may or may not be important that sufficiency was the issue here, rather than acceptability, relevance, or the availability of an external counterargument. Certainly if all the premises in a series of arguments are obviously and totally wrong, accumulation of such proofs should not move anyone closer to the other’s standpoint. But how many premises have acceptability values of zero? Some, of course, but acceptability problems seem more likely to be a matter of degree. Even low acceptability might permit some accumulation. A similar remark applies to relevance. If relevance is legitimately zero (as might happen if a standpoint is badly misanalyzed) we should not expect any rational accumulation. But a series of somewhat relevant arguments might eventually soften the impulse to refute or resist. Problems on the dialectical tier might be similarly analyzed. Conclusive objections would make accumulation irrational, but truly conclusive objections might be more common in pedagogical illustrations than in serious conversation between intelligent arguers.

An important limitation of the present study is its sample size. Collecting, transcribing, and coding conversational data is very effortful. Although the ecological validity of such a design is a good compensation, the fact remains that some of the statistical evaluations were underpowered. The key tests were based on a very modest sample size because so few of the conversations had more than five of the flawed arguments. A particular frustration is that it proved impossible to follow up on the age effect. Hamilton and Hample (in press) speculate that the first year of college is critical in acquiring arguing ability because college is the first intensive exposure to sophisticated arguments being explored open-mindedly in classrooms. With only four freshmen in the sample, however, no reasonable statistical analysis was possible. It may prove necessary to follow up on the present findings with a non-conversational design using vignettes in order to acquire a healthier sample size.
REFERENCES


APPENDIX: FLAWED CASUAL ARGUMENTS

- In Washington DC as soon as handguns were banned, the homicide rate went way up. We need to go back to what we had before, when citizens could carry guns, and then the murder rate will go down.
- New York lets citizens carry guns and Washington DC doesn’t. The murder rate in DC is about 6 times higher than the murder rate in New York. If we have New York’s handgun law, we will have less people being murdered in DC.
- Every place in the country where you have lenient gun laws, you have the lowest crime rates. So we ought to have more lenient gun laws not stricter ones.
- After Washington started its gun ban, the murder rate went up about a third when you compare the 5 years before to the 5 years after. After Chicago started its gun ban, the robbery rate went way up, too. Taking guns out of citizens’ hands stimulates violent crime.
- If citizens have their own guns, they can use them in self defense. This happens a lot in places that let people own their own handguns, like in Georgia. Not only is it good to be able to defend yourself just on general principles, but it will lower the crime rate too.
- Obviously guns are aggressive tools. There are two sides to a crime. If criminals have a lot of guns, there will be a lot of crime. But if potential victims have a lot of them, the criminals will be deterred and there will be less crime.
- After Australia put its strict guns laws into effect in 1996 violent crime went up 32% in the next 6 years. Weak gun laws make stronger criminals.
- Chicago has a gun control law like Washington’s, and Chicago also has a top ten homicide rate. That’s two out of two cities. Wherever you see a gun control law you also see a high murder rate, and you just have to wonder if that’s because citizens can’t defend themselves.
- Virginia and Maryland have weaker gun laws that Washington, but the crime rates in those states are much lower than in Washington. You really have to conclude that lenient gun laws discourage crime.
- So many people in Washington and Chicago get murdered and robbed it’s pretty obvious that they can’t defend themselves. And the gun control laws in those cities are what prevent them from defending themselves.
Commentary on “CONVERGENT CAUSAL ARGUMENTS IN CONVERSATION” by Dale Hample

KATARZYNA BUDZYNSKA

Institute of Philosophy
Cardinal Stefan Wyszynski University in Warsaw
Wóycickiego 1/3, 01-938 Warsaw
Poland
k.budzynska@uksw.edu.pl

1. INTRODUCTION

The paper discusses a very interesting topic of acceptability of convergent arguments and uses a very important (but unfortunately quite rare in argumentation theory, AT) experimental method of investigation. I must say I am very interested in this method, since I believe that such research areas as argumentation theory need to confront a theory with practice and as a result there is a need for a balance between reflection and experimental methods. AT cannot be only an armchair philosophy if it aims to be close to the natural discourse, if it wants to teach how to analyze real-life arguments. And there are a lot of ready theories that AT can use as a source and from which it can benefit – except of social science methods used in this paper, there are e.g. linguistic methods of corpus analysis (there are two corpora of analyzed arguments available online – multilingual corpus, ArgDB (Reed 2006) and Polish corpus, ArgDB-pl (Budzynska 2011)) or models of social psychology (used broadly by O’Keefe, see e.g. (O’Keefe 2007); but also in (Budzynska 2010) where Elaboration Likelihood Model is used to study persuasive arguments; or (Pasquier et al. 2006) where cognitive coherence theory is applied to build a computational model of argument in multi-agent systems). Yet, the practically-oriented approach is still under-represented in AT community. But that is enough for the ideology on methodology for argumentation theory—let us come back to the subject of the paper.

I will be rather brief in commenting on the results of the experiment presented in the paper—the numbers of the results speak for themselves. Instead I will spend some time on suggesting possible variants of the experiment which might be useful for further investigations into the issue of acceptability of convergent arguments.

2. SOME COMMENTS AND A CLARIFICATION REQUEST

Although later it becomes clear that the author concentrates on the subjective acceptability of convergent arguments, in the introduction (in the second paragraph) it seemed to me like there might be some confusion between objective and subjective levels of argument evaluation:

If all three converging arguments are poor (i.e., have serious problems either with the illative issues of acceptability, relevance, and sufficiency, or with the dialectical tier; Johnson 2000) should we conclude that the main conclusion is poorly supported, or might we consider the possibility that individually flawed reason sets might combine in some way to give more than poor support for the conclusion? (Hample 2011)
Maybe it is just unfortunate formulation, therefore I think it is worth stressing in the commentary that statistical method can provide data referring only to how people actually evaluate arguments. An experiment cannot demonstrate whether three poor convergent arguments give either poor or more than poor support for the conclusion. It can only show us whether the hearers think that three poor convergent arguments give poor or more than poor support for the conclusion.

The other issue that raises doubts was the scale used for the measurements in the experiment: “The participant’s behaviors throughout each segment were classified into one of five categories: (1) accepted the argument; (2) ignored the argument; (3) rejected the argument without giving a reason; (4) rejected the argument with explicit refutation; and (5) took up the argument and revised it to make it sounder” (Hample 2011). I agree that the last response is cognitively the most sophisticated one (which, as far as I understood, was the criterion according to which this scale was built). However, I have some doubts about the adequacy of using the last category for the hypothesis tested in the experiment. If the experiment was about to demonstrate whether people become less critical to the general conclusion (at the scale “less” means from the highest number, i.e. 5, to the lowest, i.e. 1), then (5) does not seem to me to be more critical than (4), i.e., rejecting the argument with explicit refutation. This response is in fact quite favourable with respect to the claim in this sense that the opponent concedes the claim, and only improves the argument which seemed weak to him.

Although I appreciate experimental method, it has some weakness. It can show whether a hypothesis is right or wrong, but it does not provide an explanation for the results obtained—the unintuitive results are left unexplained. For example, the experiment described in the paper showed that: “older students were more accepting of the weak arguments”. That seems quite unintuitive. It would be very interesting to find out what is the possible reason for this phenomenon. The author suggests some possible explanation, however, with the data obtain it is impossible to verify it. Similarly, the experiment “shows that as we proceed through a sequence of weak arguments, the general response becomes lower (i.e., more accepting), apart from a bump at argument position 4” (Hample 2011). I wonder if the author has any hypothesis for this result. Maybe not every weak convergent argument has the same effect on the hearers. If this is the case, then the interesting question would be: what are the differences between convergent arguments which are important for their evaluation.

In the section “Discussion”, the author says that “This investigation shows that as conversations moved through a sequence of five flawed causal arguments, people became less critical of them” (Hample 2011). This suggests that there is some kind of relation between the acceptance of subsequent convergent arguments. If this relation does not have the inferential character, then it would be interesting to examine issues like: what is the nature of this relation, whether the previous argument makes the hearer less critical about the premise of the next argument or about the conclusion (making him at some point more and more indifferent on what premises of next arguments state).
3. VARIANTS OF THE EXPERIMENT

In this section, I discuss some possible variants of the experiment described in the paper, which might be useful for the future investigation into the relation between the number of convergent arguments and their acceptability.

3.1 Processing convergent arguments in central vs. peripheral route

It would be interesting to explore what are the differences in the evaluation of convergent arguments when the hearer processes them in the central and peripheral route (Petty and Cacioppo 1986). The number of arguments serves as a peripheral cue which the hearer uses to evaluate whether he should believe the message. The experiment in (Petty and Cacioppo 1984) showed that when the peripheral route is activated the more weak arguments are given the more favourable thoughts the hearer has, while when the central route is activated the more weak arguments are given the less favourable thoughts the hearer has. This does not hold for strong arguments. Both in central and peripheral routes, the more strong arguments, the more favourable attitude towards the message. It would be interesting to see is there any differences between the dynamics of acceptability when the subsequent arguments are all convergent causal arguments and when they are of other type of convergent arguments or when they are deductive and so on. It could be also examined whether the convergent causal arguments behave differently than the arguments considered in the experiment in (Petty and Cacioppo 1984)—maybe e.g. the critical attitude decreases quicker when the arguments are convergent?

3.2 Weak vs. fallacious convergent arguments

In the paper, both weak (but not completely valueless) and fallacious arguments were used in the experiment. It would be interesting to design an experiment where only weak convergent arguments would be given and the other one when only fallacious convergent arguments would be proposed. In the peripheral route, the hypothesis is that the type of the flaw should not matter, since the receivers do not consider the content of the message. In the central route, the hypothesis would be that the fallacies would decrease acceptability faster than weak but not fallacious arguments.

3.3 Independent vs. inconsistent arguments

It would also be interesting to examine whether the dynamics of acceptability of convergent arguments is different depending on the relation between the contents of subsequent convergent arguments, like e.g.: (1) when they are completely independent (referring to unrelated data); (2) when their contents somehow overlap, e.g. showing the same data (such as “legalization of guns reduced crime...”) with small variations (such as “...reduced in New York”, “…reduced in Chicago” and so on); (3) when the contents are somehow inconsistent with each other, e.g. “legalization of guns reduced crime in most big cities in US, therefore we should make it legal in DC”, “US has one of the highest rates of crime in the world, including DC, therefore we should make guns legal in DC”.

3
3.4 Undercutter vs. rebuttal counter-arguments as a critical response

It might also be interesting to extend the scale: (1) accepted the argument; (2) ignored the argument; (3) rejected the argument without giving a reason; (4) rejected the argument with explicit refutation, with the differentiation for various types of counter-arguments such as undercutter and rebuttal. It seems like undercutting response would be more sophisticated critique and therefore could be viewed as more critical than the rebutting one.

4. CONCLUSIONS

The paper shows the results of the (important for argumentation theory) experimental method applied to investigations into interesting topic of the acceptability of convergent arguments. It also opens several exciting paths for further exploration of the deep and complicated nature of how people evaluate this type of arguments.

REFERENCES


