Enhanced Discourse Norms: Intellectual Processes

Research Report — Consensus Research, Part 3 of 6

General Facts

Research area:	Consensus	
Research topic:	Enhanced discourse norms	
Lead researchers:	Geoff Anders	
Additional help:	Many members of Leverage team	
Date of research:	I. Initial experiment	— Sep 2013
	II. Proceduralized discussion	— Early 2014[1]
	III. Continued exploration	— Sep-2013-Ongoing
Report author:	Geoff Anders	
Research outcome:	Explored, adopted	
Publication date:	Oct 2021, v1.0	

Summary

In September 2013, after previous failures, we identified an approach that very plausibly improves individuals' ability to reach consensus in a substantial number of cases. This approach involves augmenting regular discourse norms, i.e., the ways people normally engage in discussion and argument, with the option to discuss participants' intellectual processes and practices, sometimes in a very fine-grained way.

In particular, we found that in some cases, it was possible for individuals to reach agreement in cases where they plausibly would not have otherwise by adding to the conversation a discussion of the general or specific intellectual practices they employed or mental states they went through in forming their views. We found also that in many cases it was possible to examine mental actions in a much more fine-grained way than one might naively think, and that in many cases a surprisingly fine-grained examination made it easier to reach consensus.

Discussing intellectual processes and practices as a natural expansion of regular discourse continues to be a practice among a number of present and past members of the Leverage research collaboration, and judged useful by them. We welcome further tests by other researchers, and in particular would like to understand better the practical difficulties others might find in attempting to engage in discourse augmented in this way. Understanding such

[1] My memory of the general date of this experiment is hazy, and I have not discovered any documents or testimonial sources that indicate the date more precisely.

difficulties is a prerequisite to determining whether it is sensible to recommend the wider adoption of the practice.

Investigation

Background

General Background

People often seek to reach agreement through discussion, debate, and argument. This activity, which is often referred to as rational discourse (see also: dialectic), sometimes results in the participating parties, or "interlocutors," reaching agreement. And while there may be gray areas or borderline cases, there are also common canonical experiences of (1) reaching agreement with an interlocutor, and (2) failing to reach agreement with an interlocutor.

It is often valuable to have interlocutors reach agreement. As such, any method, practice, technique, or approach that helps interlocutors to reach agreement in more cases may be quite valuable, especially if that method is cost-effective, easy to learn, effective in a large number of cases, and effective in cases where other known methods fail.

It may be difficult to describe comprehensively what constitutes the "standard practice" of rational discourse. Knowledge of rational discourse is typically intuitive, with practitioners implicitly knowing what is permitted and what is not. Nevertheless, there are some explicit treatments of rational discourse, which focus on the nature and structure of arguments, or attempt to describe and enumerate the stages and rules of discourse. Apart from these more theoretical treatments, there are also basic texts on the topic for students (e.g., here and here).

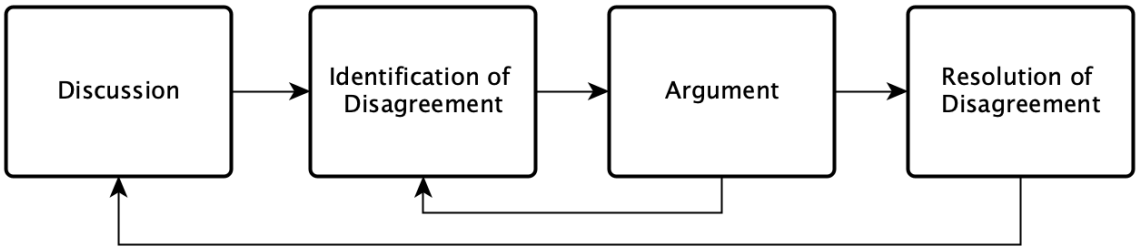
There are a number of proposed ways to enhance discourse, such as making bets or assigning probabilities to propositions. It is difficult and costly to test the efficacy of such proposals in a definitive way, but one source of evidence in favor of a given enhancement is its voluntary adoption and continued use by interlocutors who are trying to reach agreement in concrete cases.

In April 2013, I identified the examination of intellectual practices and processes as a potential way to substantially improve the ability of groups to reach consensus. We then tested a particular procedure for having a group reach consensus, through rendering participants' intellectual practices explicit and using those to navigate disagreement. This failed (Research Report), leaving us having found that people employ a large number and variety of intellectual processes and practices, but not knowing how to leverage that fact to produce practical gains.

In September 2013, I stumbled upon the fact that it was possible to describe some intellectual practices and processes in a much more granular way than we had previously tried. This led to an investigation of whether giving such descriptions could lead to substantial gains (10-100x) in helping people reach consensus. As usual, we had a strong practical interest in the results.

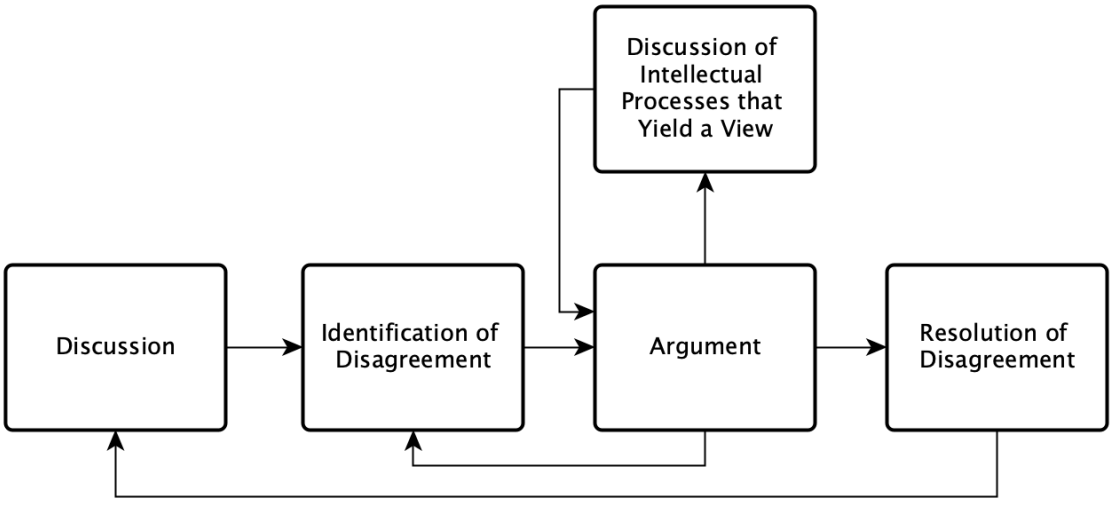
Description of Discourse Variant

A simple model of discourse is as follows: Interlocutors engage in discussion. Discussion proceeds until some disagreement is detected. Once a disagreement is detected, the interlocutors shift their focus to the identification of what precisely the disagreement is. Once they identify the disagreement, they shift to arguing about it, providing reasons and evidence. In some cases, it will become clear that they have not actually identified the disagreement, in which case they will shift back to identifying the disagreement. In other cases, they will continue arguing until they reach resolution. After resolution, the original discussion can continue. On this model, discourse fails if disagreements are not recognized, or not identified, or not resolved via argument.



Simple model of rational discourse

The proposed discourse variant adds a potential sub-conversation: when arguing but not reaching resolution, the interlocutors may choose to move to talking about the intellectual processes or practices one or the other is employing or has employed in reaching the views of their under discussion. Then, once enough information has been gathered, ideally enough to suggest new ways forward in the arguments, the interlocutors return to arguing over the disputed point.



Simple model of rational discourse, variation that permits discussion of intellectual processes

The sub-conversation devoted to intellectual processes can take a number of forms. For instance, here are two examples of a person asking an interlocutor about the interlocutor's intellectual practices:

- The person is in an argument with an interlocutor but getting stuck making progress on the discussion of the evidence. So, the person switches the topic a little and asks their interlocutor, "How did you get to that conclusion?" The interlocutor might then answer: "I thought of the most obvious thing wrong with what you were saying and said that," or "I tried to think of an example that illustrated your point, but couldn't, so I kept my opinion." The interlocutor's answer may provide the person with enough evidence to enable them to identify the actual crux of their disagreement. In such a case, the person and interlocutor might return to their argument.

- Similar scenario, except when the person asks their interlocutor, "How did you get to that conclusion?", the interlocutor might fail to answer, and instead continue giving evidence for their positions as though they were still having a regular argument. In such a case, the person may need to help their interlocutor understand the request being made of them, and may need to walk them through paying attention to their mental states and giving answers that correspond to them.

It is also possible for a person to simply report on their own mental processes as part of giving greater context for their statements.

- For instance, a person might say "my mental simulation yields that John will say no to that proposal" or "my gut reaction is that John will say no to that proposal," or "my initial take after a few seconds is that John will say no to that proposal," or "I tried to think of a plausible scenario on which John would say yes but didn't think of one, so I think he'll say no," etc.

- Similarly scenario, except the person gives inaccurate reports of their own mental processes, and thus must be prompted to look again to see if it was possible to give a more accurate report. For instance, an interlocutor might say, "My best story is that John will say no to the proposal," but then the person will prompt them to recognize that they actually are giving (e.g.) the first story that popped into mind, rather than their best story.

Part of the art of doing this properly is being able to precisely describe one's mental practices, to help others to be able to do the same, and to know what to do in an argument in response to a given set of mental practices. In some cases, this can involve an extremely granular examination of intellectual processes.

- For instance, a person might ask their interlocutor, "What mental process did you use to conclude that Susan will call tonight?" The interlocutor might then answer, "I don't know, it just occurred to me." The person may then lead the interlocutor through noticing that when they pay attention to whether Susan will call, they first imagine Susan not calling, then quickly imagine the most plausible context in which that makes sense, and then feel a feeling associated with mismatching mental content or inelegance. Having found this,

the person and interlocutor may then return to the argument, where the person may not start by searching for plausible scenarios according to which Susan will not call.

Approach

Our investigation of using discussion of intellectual processes and practices to aid regular rational discourse had three parts: (1) an initial experiment where I tested the idea, (2) an experiment soon after, where I attempted to routinize the reporting of intellectual processes, and (3) many years of informal testing by approximately 10-20 researchers.

In the initial experiment, I waited until I had a discussion aimed at reaching rational agreement but where all of the normal paths forward I would normally take had failed. Then, rather than continuing to discuss my and my interlocutor's reasons and evidence for our views, which we had already tried, I asked them to describe how their views had come about, and what mental processes they used in forming those views. As they answered, I asked further questions to build a mental model of how they had come to their current position. I then used that information to identify the source of disagreement. Finally, I attempted to resolve the disagreement using the standard means employed in regular rational discourse. I estimated whether the disagreement was resolved using the means normally employed in rational discourse.

Some time after this initial experiment, I tried to conduct a discussion where I and my interlocutor were required to state the intellectual methods we used to generate the next thing we said in the conversation. The conversation thus went:

(1) Interlocutor A makes statement #1
(2) Interlocutor A explains the mental processes that led to statement #1
(3) Interlocutor B makes statement #2
(4) Interlocutor B explains the mental processes that led to statement #2
(5) Interlocutor A makes statement #3
(6) Interlocutor A explains the mental processes that led to statement #3

…, etc.

Beyond these two explicit experiments, and starting after the first one, I and other researchers continued to supplement regular discussion (at appropriate times) with discussion of the intellectual processes and practices the interlocutors were employing. Over the long run, the primary questions of interest were first, whether researchers continued to use this discourse variation voluntarily over a substantial period of time, even in an environment where there were a large number of alternative practices, techniques, methods, and so forth competing for their interest that they could use instead, and second, whether researchers endorsed this continued use.

For the preceding research, we did not consult external literature. The initial experiment took place in Sep 2013, the second experiment took place in early 2014, and the ongoing use of the discourse variation started in Sep 2013 and continues today.

What We Found

Our findings are stated in the following four sections:

1. Discourse with Intellectual Processes
2. Attempted Routinization
3. Long Term Use Patterns
4. Best Explanation of the Evidence

In the first two sections I will cover what we found. Each section is brief. In the third section I will provide general evidence about the long term use patterns of the discourse variation, while giving examples. The fourth section is devoted to giving our analysis of the data from the preceding sections.

Discourse with Intellectual Processes

The first experiment was brief, and I did not keep a record of the specific content. After having been otherwise stymied in the conversation, I asked about the intellectual processes that my interlocutor had used in coming to their view. They answered, and this enabled me to pinpoint a crucial source of disagreement. We then discussed that source of disagreement, resolved it, and reached agreement, judging in the way typically employed in rational discourse.

Attempted Routinization

The second experiment was brief as well, and I also did not keep a record of the content. We tried to follow the strictures I had described. After a few back-and-forths, my interlocutor and I agreed this was both cumbersome and not useful, and gave up.

Long Term Use Patterns

From 2013 through 2019 and after, many of the people who learned how to think about people's various intellectual practices adopted and continued using the discourse variant under discussion. That is to say, it became for many people a standard conversational option to ask about, or oneself report on, a person's intellectual practices or processes. People used it on a wide variety of topics, though not so frequently as to be ubiquitous. A rough estimate might be that 10-15 people learned to use it well enough for it to count, they used it in an average of 5-30% of arguments, and that most of them continued to use it through mid-2019. When used, the discourse variant typically did not take much time.

Further, in addition to continued use, we encountered no reports that people who continued using the discourse variant were displeased about their continued use. It just became another normal part of conversation, a useful tool to be employed in some circumstances.

Best Explanation of the Evidence

The best explanation we reached with respect to the discourse variant, with respect to effectiveness, cost-effectiveness, etc., is:

(1) **Detectable when effective.** Researchers have reference experiences of what it is like to reach agreement or to fail to reach agreement in an argument. They could then use that reference experience to identify whether they were reaching agreement in the cases where the discourse variant was being employed. This makes it unlikely that they were incorrect about its effectiveness.

(2) **Effective in many cases.** The researchers were generally very motivated to reach agreement in many cases. Their use of it in (very roughly) 5-30% of cases indicates that it applies to many cases in terms of absolute numbers.

(3) **Sometimes effective when other approaches fail.** The discourse variant not infrequently works in cases where no other plans have worked. It is difficult however to estimate the true magnitude of this.

(4) **Cost-effective to use.** The fact that the discourse variant does not take that much time to employ, combined with its effectiveness in practice, implies that it is rather cost-effective, at least once you are able to use it well.

(5) **Somewhat difficult to learn.** The use pattern does not indicate great ease of use; if this discourse variant was very easy to use, it would have been adopted by many more people.

This best explanation is tentative; a more thorough examination of the data could yield changes.

Conclusions

Main Conclusion

Our investigation yielded one primary conclusion, which is that it is often useful in rational discourse to have the option available to have a sub-conversation about the intellectual processes or practices that were used or are being used as part of the person forming and retaining their beliefs. It appears effective in many cases, including some where other known practically available approaches fail. It also appears cost-effective, though it may be difficult to learn.

While it appears to be effective for helping increase people's ability to reach consensus, it is worth noting that it by itself does not produce a 10-100x increase. Any intervention that improves each person's ability to reach consensus may yield superlinear returns. But with that said, and speaking roughly, it seems unlikely that this by itself would yield more than 2-5x improvement.

Conditions for Revisiting

We would revisit this if we encountered reason to think that many of the researchers who learned this approach stopped using it, or if they stopped endorsing its usefulness. That might mean a number of things, but it might indicate a perception of a lack of or diminishing usefulness.

We would also revisit this if we learned that people's subjective experiences of reaching agreement or failing to reach agreement with people were substantially misleading.

Next Steps

Several questions remain on the topic of enhancing discourse norms through the examination of intellectual processes. These include:

- Are there cost-effective ways to more accurately measure the effectiveness of interventions like this one? What about this intervention (i.e., the discourse variant) itself?
- Are there particular easily identifiable classes of disagreement that discussion of intellectual processes would be especially helpful for? What are the signs of those classes?

After we got started on intellectual processes, it is unsurprising that we soon started trying to experiment with other potential additions to the standard set of discourse norms. The next modification we found, and the next major line of approach we took to trying to resolve the problem of consensus, had to do with the psychological inputs to beliefs.

- Research Report on Enhanced Discourse Practices: Psychology *[forthcoming]*

Retrospective

Oct 2021:

It has been slightly more than two years since the end of the Leverage research collaboration. In that time, it has been harder to keep in touch with former staff and collaborations. For the people with whom I am in touch, and who gained facility with the relevant practices, the activity of pointing out one's own mental moves or asking at times about the mental moves of others remains commonplace.