

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/320616866>

# Do reasoning limitations undermine discourse?

Article in *Thinking and Reasoning* · October 2017

CITATIONS

3

READS

408

2 authors:



**Deanna Kuhn**

Columbia University

175 PUBLICATIONS 12,667 CITATIONS

[SEE PROFILE](#)



**Anahid Modrek**

University of Pennsylvania

8 PUBLICATIONS 18 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Cognitive regulation outdoes behavior regulation in predicting state standardized test scores over time [View project](#)



Can autonomy engender effective reasoning? [View project](#)



## Do reasoning limitations undermine discourse?

Deanna Kuhn<sup>a</sup> and Anahid Modrek<sup>b</sup>

<sup>a</sup>Department of Human Development, Teachers College, Columbia University, New York City, NY, USA; <sup>b</sup>Department of Education & Department of Psychology, UCLA, Los Angeles, CA, USA

### ABSTRACT

Why does discourse so often seem shallow, with people arguing past one another more than with one another? Might contributing causes be individual and logical rather than only dialogical? We consider here whether there exist errors in reasoning that could be particularly damaging in their effects on argumentative discourse. In particular, we examine implications for discourse of two such errors – explanation as a replacement for evidence and neglecting the likelihood of multiple causes contributing to an outcome. In Studies 1 and 2, we show these errors to be prevalent in a cross section of adults, as well as samples of community college students and young adolescents, with minimal age-related improvement. They also occur, although less frequently, among a sample of highly educated adults, and in Study 3, we examine their role in the discourse of college-educated adults. We point finally to evidence that these individual reasoning errors are potentially addressable through education.

**ARTICLE HISTORY** Received 9 November 2016; Accepted 2 October 2017

**KEYWORDS** Reasoning; discourse; causality; development; education

Reasoned discourse, long regarded as the lifeblood of democratic societies, appears more endangered than ever before by polarisation, insularity and communication confined to sound bite and slogan. Are cultural factors to blame, or is it simply human nature that's at fault? People are thought to be "cognitive misers" (Stanovich, 2009, 2010) operating most often under a least-effort principle, with intellectual values too weak to support the effort that thinking deeply requires. Furthermore, they may lack the foundation of mature epistemological understanding that reasoning and discourse require, as a result failing to see their purpose and hence to value the effort they entail (Moshman, 2015; Kuhn, 2009; Ricco, 2015; Richter & Schmid, 2010). If knowledge consists of claims not open to question – either facts that can simply be "looked up" or opinions to be accepted without scrutiny as private possessions of their holder (the stances reflected in less mature epistemological positions), there is little purpose to discourse.

**CONTACT** Deanna Kuhn  [dk100@tc.columbia.edu](mailto:dk100@tc.columbia.edu)

© 2017 Informa UK Limited, trading as Taylor & Francis Group

In addition, people may fail to appreciate essential norms regarding the conduct of discourse (Van Eemeren, Garssen, & Meuffels, 2009; Walton, 2014). Thus, people may use argument only as a device to further their own objectives (Mercier & Sperber, 2011). Or they are prone to “belief overkill” (Jervis, 2006; Tetlock, Vieider, Oatil, & Grant, 2013) and don’t bother, rarely giving conscious attention to their beliefs or seeking to test them against reality, although their actions are belief and expectation driven.

Here we pursue a possibility raised by Rips (1998), asking whether, in addition, there may exist errors in the reasoning itself that individuals employ that would be particularly damaging in their effects on argumentative discourse and even contribute to its seeming decline in quality. In other words, could threats to discourse be individual and logical rather than only dialogical? In contrast to an argument as an individual product (where a typology of logical fallacies has long been identified by philosophers and psychologists), our question is addressed to the effects of such errors on the interpersonal process of argumentation, a process in which two or more people jointly address a controversial topic.

Without claiming them as exhaustive, we identify here two types of individual reasoning that people may bring to argumentative discourse that we allege seriously threaten its productivity. One has received some attention in argumentation literature but the other hardly any. Both can be regarded as competence errors, since they occur in contexts in which respondents have been asked to reason to the best of their ability. Both pertain to what is one of the most frequent topics of both informal and scientific discourse – what causes a particular phenomenon to occur? Specifically, they involve the forms of evidence that strengthen or weaken causal inferences.

The first is the substitution of explanation for evidence. When asked for evidence to support a causal claim, respondents of all ages often offer an explanation of how the alleged cause *could* produce the outcome (a mechanism account), rather than any evidence that the cause *does* produce the outcome (Brem & Rips, 2000; Kuhn, 1991, 2001; Koslowski, 1996; Ahn, Kalish, Medin, & Gelman, 1995). In effect, the plausibility of the explanation serves as the evidence, with the further effect of drawing attention away from any quest for genuine evidence.<sup>1</sup> The effect can be more damaging to the extent explanations are rehearsed and investment in them forms, further disinclining one to seek or be responsive to evidence (Kuhn & Katz, 2009; Williams, Lombrozo, & Rehder, 2013; Rittle-Johnson & Loehr, 2016). This tendency can be particularly damaging to argumentation. In a word, it allows one person to claim “Here’s how it happens,” and the other to counter, “No, here’s how it happens.” Such an exchange leads quickly to a dead end, and most limiting,

---

<sup>1</sup>As Brem and Rips (2000) point out, however, people may turn to explanation as a substitute when they perceive evidence to be scarce or absent.

one in which neither has examined the other's claim. If the arguers do contemplate one another's claims, the grounds for comparison are limited to ones of plausibility.

Second, and at least as limiting in its effect on argumentation, is an inadequate mental model of causality, specifically one that confines itself to a single cause as sufficiently accounting for an event. To illustrate first in a physical domain, both children and adults tend to regard only a single (more dominant) force to determine the outcome when two physical forces act upon an object (Goksun, George, Hirsh-Pasek, & Golinkoff, 2013). This error extends to social domains and can be connected to the long-standing social psychology concept of discounting (Kelley, 1973) – attribution of an actor's behaviour to one cause makes attribution to another cause less likely – but its status as a more general cognitive characteristic, and potential liability, has not been fully probed, particularly with respect to argumentation (see, however, Goedert, Harsch, & Spellman, 2005; Laux, Goedert, & Markman, 2010).

Single-cause thinking is clearly simplistic thinking. Real-world outcomes are most often the consequence not of a single cause but of multiple factors acting in concert, a fact that practising scientists are well aware of and take into account in both their theoretical models and empirical investigations (Sloman, 2005). Not recognising that multiple causes typically contribute to an outcome is arguably severely limiting to thinking in general, as well as discourse. If a single cause is regarded as sufficient to bear the explanatory burden, alternative causes will be seen as contradictory: Either my cause or your cause must be the correct one. An affective component enters in and reasoning becomes motivated by allegiance to one's preferred cause, with the alternative cause seen as threatening to replace it, when in fact it may be unnecessary to choose between the two. (In contrast, in the case of the first error described above, the explanation-as-evidence error, two mechanism explanations may be contradictory.) Do we, for example, look inside the abuser or outside the abuser to find the cause of drug abuse? A single-cause thinker will never develop a very deep understanding of the phenomenon. Nor is he or she likely to sustain a rich conversation on the topic.

To document these two errors – explanation as a replacement for evidence and single instead of multiple causes – in a hypothetical discourse context, in Study 1 we asked participants to select which of three forms of evidence would be most powerful as a counterargument to an opponent's causal claim. Participants were from two groups – (a) a cross section of average adults, and (b) middle-school students from an urban public school serving a diverse population. We included the younger group to ascertain whether any development in the forms of reasoning assessed was evident between early adolescence and adulthood.

## Study 1

### Method

#### Participants

For the purpose of securing as close to a random sample as feasible of the adult population, adult participants were solicited from occupants of the waiting room of a large urban train station. Very few declined. They completed the written task independently, answered several demographic questions and received \$5.00 as token compensation. Participants required on average 10 minutes to complete the task. Of the 41 participants included in the adult sample, 59% were female; 32% were between ages 18 and 25, 27% were between 26 and 45, 29% were between 46 and 65, and 12% over 65.

The younger sample consisted of 62 middle-school students (grades 6 and 7; age range 11–13; 33 female) from a public school in a working to middle-class neighbourhood of the same city. The school population was 52% Caucasian, 15% Asian, 5% African-American, 4% Hispanic and 24% of mixed background. Approximately 10% qualified for free or reduced-price lunch. Students completed the task in their classrooms, requiring on average 15 minutes to do so.

#### Task

The three task questions appear in [Table 1](#). Each of the questions posed entailed a choice among three options. (Both questions and response options appeared in counterbalanced order.)

For all three questions, option A introduces no evidence and simply makes another causal assertion. It thus does not directly address the initial claim, as a result not serving to address the stated objective of showing this claim to

**Table 1.** Task questions.

- 
1. Some health officials have found cancer rates higher in cities than in outer areas. Dr J. Rawls claimed tanning salons are to blame. Circle ONE piece of evidence that would be best to use if you wanted to argue he was wrong.
    - A. Air pollution is a more likely cause of cancer in the city.
    - B. Many people who don't go to tanning salons also get cancer
    - C. Many people outside the city also go to tanning salons and don't get cancer.
  2. People from some countries have longer average life expectancy than people in others. Dr. F. Cole claimed a diet high in fish causes long life. Circle ONE piece of evidence that would be best to use if you wanted to argue he was wrong.
    - A. Exercise is a more important cause of long life.
    - B. People who don't eat fish often live to an old age.
    - C. People who eat a lot of fish often live only to an average age.
  3. Venezuela is a country with money trouble, unable to pay its bills. Dr P. Garett claimed the cause was too many social programs to help people. Circle ONE piece of evidence that would be best to use if you wanted to argue he was wrong.
    - A. Poor money management is a more likely cause of a country's money trouble.
    - B. Some countries like Haiti have very few programs to help their people and Haiti has serious money shortages.
    - C. Some countries like Sweden have many social programs and are not in money trouble.
-

be wrong. Unless one subscribes to the mental model that only a single cause produces an effect, identification of the second cause does not bear on the validity of the initial cause.

Options B and C both introduce evidence, but of different forms. Option B states evidence of an alternative sufficient cause, i.e., the outcome may appear in the absence of the alleged cause due to another cause being sufficient to produce it. It thus does not counter the claim that the alleged factor is a cause. Option C, in contrast, does directly counter the claim that the alleged factor is a cause in that the evidence it presents is of a failure of the alleged cause to produce the outcome.

Scientifically trained reasoners would of course be within their rights to reject all three response options as insufficient (on qualitative and/or quantitative grounds), to permit any conclusion. Skilled statistical reasoners would regard the evidence as inconclusive unless frequencies were available for all four cells (presence and absence of cause and presence and absence of outcome); however, untrained individuals, we know, rarely consider more than one or at most two of these cells (Schustack & Sternberg, 1981). None of our participants expressed such uncertainty; all read carefully and contemplated the three options and chose one of them.

## Results

In Table 2 are summarised the response patterns of the two groups. Classified in Table 2 are respondents who showed a dominant response type, i.e., who chose one of the three evidence types for two of the three or all three of the items. (Those classified as showing no dominant response type chose each evidence type once.)

Considering only participants who showed a preference for one of the three evidence types, over half (56%) of the adult participants chose a counter assertion (option A) as the strongest evidence against the original claim, and less than a quarter (24%) the correct option C of evidence of an ineffective cause given the failure of the cause to produce the outcome. Among adults, education (but not age group or gender) was associated with preference. High school graduates most often preferred option A, while option C was favoured mainly by those with post-college education (see Table 3) The younger

**Table 2.** Distribution of response types in two samples.

Dominant response type	Frequencies and proportions showing	
	Early adolescents	Adults
A. Counter-assertion without evidence	21 (27.3%)	23 (46.0%)
B. Evidence of alternative sufficient cause	25 (32.5%)	8 (16.0%)
C. Evidence that cause ineffective	16 (20.8%)	10 (20.0%)
(No dominant response type)	15 (19.5%)	9 (18.0%)

**Table 3.** Response types by education level in adult sample.

Dominant response type	High school graduate	College graduate	Post-graduate degree
A. Counter-assertion without evidence	75%	58%	31%
B. Evidence of alternative cause	19%	25%	15%
C. Evidence cause is ineffective	6%	17%	54%
[Total sample]	[39%]	[29%]	[32%]

Note: An option C preference is significantly more likely among the college educated group,  $\chi^2 = 4.68$ ,  $df = 1$ ,  $p = 0.030$ .

sample did not show a significantly different proportion of correct choices (option C) than the adult sample – 20.8% vs. 20.0%. Hence there is no indication of age-related gain in this regard between early adolescence and adulthood. If anything, adults were even more likely than adolescents to favour the arguably least correct option A, which ignores the hypothetical opponent’s alleged cause entirely ( $\chi^2 = 4.98$ ,  $df = 1$ ,  $p = 0.026$ ). (It should be noted, however, that the hypothesis cannot be rejected that the younger group showed no preference among the three options.<sup>2</sup>)

## Discussion

A three-item, closed-ended questionnaire does not, by prevailing psychometric standards, provide a precise or rich assessment of an individual’s competencies or dispositions. We thought it ill-advised to include a larger number of items for two different reasons – fatigue and boredom in the younger group and time requirement in the older one. Nonetheless, at a group level, the results allow us to conclude that neither group displays a modal level reflecting mastery of the causal reasoning competence of concern to us here.

These results, however, leave us with a number of unanswered questions. Foremost, how accurate is our short, three-item assessment? Is it possible that it underestimates competence? Do a third to half of typical adolescents and adults truly regard an alternative explanation as the strongest counterargument to a causal claim? One possibility that needs to be considered is that these individuals mistakenly assume the alternative-cause response option A as a claim supported by its own (unstated) evidence, given it is offered as a potential counterargument and is identified in the question as a potential “piece of evidence”. In Study 2, we addressed this possibility by avoiding reference to this potential counterargument as evidence and referring to it simply as a statement, thus asking the respondent to “Choose the best statement to use if you wanted to argue that this person was wrong”.

<sup>2</sup>Evidence against the interpretation of random responding, however, exists in the form of an inductive reasoning task administered to the younger Study 1 participants on a separate occasion, as part of other research. The relation between performance on that task and the one used in Study 1 was significant (with those favoring Option C more likely to perform well on the inductive reasoning task), making it unlikely that participants were choosing randomly in the case of the present task.

A second objective of Study 2 was to better establish the accuracy of the three-item instrument by comparing it to performance on an instrument identical in form but containing more items. In so doing, we reduce the likelihood of classifying as a reasoning preference what is no more than a chance choice of the same option type over multiple items. We also anticipated that in so doing we might reduce the proportion of respondents who showed no preference, since a larger number of items make it more likely that such a preference will emerge if one exists.

Third, we wished to further examine and document the notable finding of an absence of developmental advancement between early adolescence and middle adulthood in the competencies examined.

## Study 2

Study 2 was conducted to fulfil the three objectives just indicated. The task was identical to that used in Study 1 with two exceptions. First, it contained eight instead of three items (see the Appendix for full set of items), identical in form to the Study 1 items and of varied content, three of the eight the same three items used in Study 1. Second, in the question posed, the term “evidence” was replaced by the term “statement” for the reason indicated above.

### *Method*

#### *Participants*

The 36 adult participants were students of equally represented genders enrolled in a psychology class at a public suburban community college in the Northeast United States. All but one were between the ages of 18 and 28. About half attended full-time and half part-time. Roughly half of students at the school are Caucasian and most of the remainder equally divided between African-American and Hispanic.

The 54 young adolescent participants came from two Grade 8 classrooms of an urban public school in the Western United States. All were 13 or 14 years old and of equally represented genders. Their ethnicity was 75% Hispanic, with the remainder Caucasian, Asian or of mixed race. Sixty per cent of students qualified for free or reduced-price lunch. The sample was thus slightly older but more economically disadvantaged than the Study 1 younger sample.

#### *Procedure*

The eight-item task was presented to each sample in a classroom setting and completed individually in writing. Three of the eight were the same items used in Study 1 and the remaining five were identical in form (see the



Appendix for full set of items). Items appeared in a counterbalanced order across participants (as did response options within items). Younger participants required somewhat longer, but all participants completed the task within 15 minutes.

## Results and discussion

Results for the two groups appear in Table 4. Each respondent was classified into one of the categories shown. The final row contains those who showed no dominant pattern, choosing each of the three options A, B and C approximately equally often. (Frequencies by individual items appear in Appendix 1.)

These percentages are similar to those observed in Study 1, using the shorter instrument, in particular with respect to the proportions showing the correct option C as the dominant response type. Also, in both studies, these proportions vary little by age (although the Study 2 adults are slightly less likely and the Study 2 teens more likely to show the least advanced option A as the dominant response, neither difference reaching statistical significance,  $p > 0.16$ ).

Thus, it can be concluded that the shortness of the Study 1 instrument was not a cause of the failure of a preference to emerge among some Study 1 participants. Instead, it appears that near one-fifth of respondents of all ages genuinely show no preference, most likely because they process the items only superficially or are entirely driven by the item content.

The fact that the proportion of participants preferring the A option as their most frequent choice is slightly lower among the Study 2 college students than among the general population of adults in Study 1 makes it possible that the word change across studies thus could have contributed to Study 2 adults doing better in this respect. The difference across studies, however, was slight, indicating that this was not a major factor.

**Table 4.** Distribution of response types across groups.

Dominant response type	Frequencies and proportions showing	
	Early adolescents	Adults
A. Counter-assertion without evidence	24 (44%)	10 (28%)
B1. Evidence of alternative sufficient cause, but acceptance of A	6 (11%)	7 (19%)
B2. Evidence of alternative sufficient cause, and rejection of A	5 (9%)	9 (25%)
C. Evidence that cause ineffective	9 (17%)	4 (11%)
N. No dominant type	10 (19%)	6 (17%)
(Total)	54	36

Note: The criterion for dominant type was choice of that type for 4 or more of the 8 items, with choices for the remaining 4 items distributed across the other two major categories. The criterion for B2 was 0 or 1 choices of A. Over the 8 items, 50% of teens and 42% of adults showed only a weak response-type preference (the preferred type chosen on 4 of the 8 items). Showing a stronger preference (5 or 6 of 8 items) were 42% of adults and 32% of teens. No participants showed an extremely strong preference (7 or 8 of 8 items), leading to the conclusion that content variation continues to be influential across the expanded item pool.

The larger number of items in the Study 2 instrument also allowed us to differentiate the group showing a preference for the B option into two subgroups. One subgroup, while most often choosing B, continued to accept the A option as the best choice on two or more items. The other subgroup, also most often choosing B, tended to reject A, choosing it no more than once.

Participants in the C-preference category also were most likely to reject A, choosing it no more than once (all adults and the majority of adolescents in the C category did so). The existence of these subgroups is suggestive of a developmental progression: individuals first recognise the strength of genuine evidence by beginning to choose the B option (without recognising the superiority of C), but they continue to endorse the A option as constituting evidence. Only subsequently do they (a) reject the A option, and (b) recognise the superiority of C.

Finally, to evaluate the power of the short Study 1 instrument, we examined the responses of the Study 2 adult sample on the three items identical to the three that constituted the Study 1 instrument. We identified the classification each Study 2 participant would have received based on just the three Study 1 items and compared it to their classification based on the larger number of items in Study 2. Among the 30 Study 2 adults who showed a preference for one item type, that preference failed to emerge based on only the three Study 1 items for only a minority of participants – 7 of the 30 (23%) – allowing us to conclude that the short Study 1 instrument does a serviceable job of identifying such preferences when they exist.

### Study 3

The findings from Studies 1 and 2 showed adults to be little (if at all) more advanced than young teens with respect to the competencies being examined. Before accepting this conclusion, we questioned whether a group that was overall more accomplished intellectually would be more likely to subscribe to rigorous standards of evidence for countering a causal claim. The association with education observed in Study 1 gave us reason to think so, and we therefore pursued this possibility in Study 3. In addition, in Study 3, we sought indication that the individual differences identified in Studies 1 and 2 would in fact be detectable in such individuals' discourse and have identifiable consequences.

The uniformly high intellectual level of the Study 3 sample is an advantage in pursuing our second goal of examining consequences regarding their discourse, since other possible sources of variance are reduced. We include a qualitative examination of dialogues by two pairs of adults selected from this and a comparison sample.

## Method

### Participants

A sample that fulfilled our objectives consisted of 81 students (38 female) in an MBA degree programme at a major Ivy League university and enrolled in an advanced elective course devoted to strategic decision-making. A large part of the course consisted of participating in a simulation of a business environment. Students entering the MBA programme almost all have prior experience in positions in the business or non-profit world, and in one of the two class sections students continued their professional employment and attended the programme part-time. All held at least bachelor's degrees and many had earned other post-graduate degrees prior to entering the MBA programme. Their ages ranged from mid-20s to early 40s.

### Procedure

As part of a self- and course-assessment during the final class session, Study 3 students completed the same questionnaire administered in Study 1. It was described to them as an assessment of their answer to the question, "What, in general, do you consider good evidence?" It was included as the final question, along with other questions more directly associated with the course.

Previously, as the focus of the course, groups of 3–4 students randomly formed in advance by the instructor worked together as a firm to compete against other firms in the industry. One section was larger than the other and contained 15 groups. The other section contained 10 groups. Eight groups were composed of four persons and 17 groups had three persons.

Each group's objective was to maximise firm profits relative to other firms over a multiple-year time period. Over the course of the simulation, a group had to collaborate and come to agreement as a group in making a series of strategic decisions involving multiple variables potentially affecting the firm's performance, thus making the task a fitting one in terms of drawing on the reasoning competencies of concern to us here. Following each set of decisions they received feedback on multiple factors, feedback they then needed to evaluate and take into consideration in making their next set of decisions.

Groups spent approximately 24 hours collaborating within their group in eight cycles of reaching decisions, analysing interim outcomes, and making new decisions, over the course of the class. The time allocated for each group discussion ranged from 2.5 to 3.5 hours, and groups typically continued their discussions for the full period. In each one, the group had to examine the newly available data (e.g., their firm's revenue, costs, inventory, and profit; competitors' performance; market research results; market forecasts) and reach consensus as to how to proceed. Thus, drawing on evidence to reason about how multiple variables influence outcomes was clearly required and this reasoning needed to be collaborative, allowing the group to reach joint decisions.

## Results

### *Descriptive findings*

The large majority of the 83 participants in the Study 3 sample showed a high level of performance on our assessment. The percentage selecting Option C – evidence the cause is ineffective – as the strongest evidence against a claim as a dominant or consistent response was 78%, higher than that of either the Study 1 sample or any of the Study 1 education subgroups (Table 3) and higher than the Study 2 college student group.

There nonetheless continued to exist individual variation. An Option B (evidence of alternative cause) dominant pattern was made by 14% of participants, and an Option A (counter-assertion without evidence) dominant pattern was made by 7%. Furthermore, in contrast to the groups in Studies 1 and 2, there occurred no participants who failed to show a dominant response pattern, i.e., failed to show a preference for one of the three options across two or more of the three items, and 39% of participants showed a consistent preference across all three items (i.e., answered all three in the same way).

### *Qualitative analysis*

We wished to first study qualitatively representative dialogues of individuals who showed contrasting patterns on our assessment instrument, to see if we would detect any contrasting characteristics in their patterns of discourse, as this was a major goal motivating our investigation. For this purpose we chose two individuals from Study 3 who had shown a consistent “C (evidence cause is ineffective)” pattern on our instrument and asked them to engage in a dialogue with one another. Since no Study 3 participants showed a consistent A pattern, we solicited two volunteers who had shown this pattern from an additional sample of adult students enrolled in graduate-level professional development training to whom we had also administered our instrument. Both were public school teachers. The two A-pattern volunteers and the two C-pattern volunteers were thus all college graduates and all had done some graduate work, although only the C-pattern volunteers had attained a graduate degree.

We asked each pair to choose from a list of suggested topics one on which they held opposing views and to engage in a dialogue on it. They were asked to try to reach agreement if they were able to. Their two dialogues are presented in Tables 5 and 6. It warrants emphasis that we did not collect such dialogues from a larger number of volunteers and choose from among them the ones presented here. Rather, we invited only one pair among those who had shown the respective patterns of interest to volunteer; both accepted and their dialogues are the only ones collected and the ones shown here. Both pairs conducted their dialogues in private, recorded them and made the recordings available to us.

**Table 5.** Dialogue between two C-pattern Individuals.

Topic: Should smoking be reduced by educating people about its dangers or by charging a very high tax on purchase of cigarettes?

P: I favour education. Smoking is a personal decision. Something intrinsically very addictive and something people need to understand and make a decision for themselves. While I understand that people might vote, might purchase based off of their pocketbooks, you have to pay for smoking and if people really want something they're gonna find out how to do it probably to the detriment of other areas where they could be spending some of that disposable income.

N: I'm taking the other position that there should be a tax. There's plenty of evidence to suggest that smoking kills and the government has a responsibility to stop people hurting themselves.

P: I agree the government has a responsibility to stop people. I think we just disagree on the means by which they do this. And I'm going to point to two data points that I think rebut and actually state that raising taxes and making people decide based off of their pocketbooks has not been effective. I think the first thing we can talk to are a number of illegal drugs right now that are on the street. You see people who have very little money don't purchase food but they find the means to buy those drugs by any way possible. By the fact that there is a high price they're not only going to be purchasing them, to their detriment they're not going to be purchasing the things they need That's my first argument.

N: Let me disagree with that. You have a point that people do buy illegal drugs, But on the other hand the government has a responsibility, and there are many areas where governments do take action. to help people. Drugs is certainly one. There are a lot of other products people cannot buy because the government thinks it's bad either for them personally or for other people. And the fact that people are getting illegal drugs I think does not stop government's responsibility for trying to stop people from smoking by putting a high tax.

P: I don't think we disagree about whether it's the government's responsibility. It's the means by which they do it. I don't disagree it's the government's responsibility to educate, put programs in place. But I think the government should allocate those resources to education, not taxes.

N: I think people should be forced to pay. I think they should ban cigarettes altogether. But failing that. by making it really expensive to people is a good second best.

P: But if you had to pick one or the other, and the objective is to stop people from smoking. I believe it's a combination of the two. But if you had to pick one, is it higher taxes or education? And I think there's a lot of evidence...and I'm going to point to Denmark where I was watching a documentary where they actually legalised and kept the price the same – this was for some hard drugs –when they legalised it and they continued to educate the people – I don't have the data in front of me – but the amount of usage was reduced. This is one case study which might be contrary to the argument for raising taxes.

N: People have been educated about the dangers of smoking for years. You even have to put on the cigarette box how dangerous it is to smoke. So it's pretty clear that doesn't happen. On the other hand, people do get worried about their pocketbooks and what they pay and I think that a higher price they have to pay will probably reduce their ability to smoke. There's probably been studies on that of when taxes have gone up in the past. I don't have that data in front of me but that would be something worth looking at.

P: I would tend to argue that between the 70s and 2016, if you were to look at the contributing factors, there's been a huge decrease in the rate of smoking in the last 30 to 40 years, as a per cent of population between the late 70s and 2016. If you were to try to dissect the factors that impacted that, you might find that in areas where there was a high tax, really there wasn't a decrease in smoking. So there's really no corollary\* between a high tax and a decrease. But also schools that really focused on educating people, when in fact there was no increase in tax, you would find a decrease in smoking.

N: I'm sure there's data there and I think you're right, smoking has gone down over the years. But I think you have to look at the data and tease out of that data whether it was education or whether it was taxes. And I believe you will find that taxes had much greater effect than the level of education.

\*[correlation]

The two characteristics that our instrument was designed to assess are both evident in the dialogue between P and N. Both P and N refer to actual or potential evidence as the essential basis on which a claim is supported (e.g., "I'm going to point to two data points" and "I'm sure there's data there").

**Table 6.** Dialogue between two A-pattern Individuals.

---

Topic: Is the cause of teacher turnover low pay or poor working conditions?

A: So I think teachers are treated poorly for the amount of work they have to put in.

O: Maybe for some, but at the end of the day if salary was higher more teachers would probably stick around.

A: Not sure if I agree; it's how people treat you.

O: But you have to admit money incentivises most people.

A: I think how you feel when you come to work and how appreciated you are is a stronger incentive.

O: So money has nothing to do with how happy or appreciated teachers feel?

A: I think working conditions, like administration and support, has a stronger impact on how we feel.

O: But salary would at least make more teachers stay.

A: Okay, teachers don't work for pay.

O: I didn't say that. I just think that higher salary would change the turnover rate.

A: Not sure if I agree; I mean, think of that lack of support from administrators.

O: Well there is need for more support from everyone

A: Well yeah.

O: But turnover is high because many realise they aren't compensated enough for the amount of work they do.

A: Teachers do not get into this field because of wages.

O: We're asked to do many other things besides just to instruct in the classroom and many are hardly making ends meet with the amount they get paid.

A: Okay, fine, but the reason for turnover is the way schools are run, not the money.

O: Salary change would make people want to stay.

A: Teachers go into the profession with a general idea of the salary but they can't predict the work conditions.

O: Not everyone knows what they're getting into.

A: I mean, I know.

---

Second, both understand that the two factors under discussion as potentially producing a desired outcome are not mutually exclusive alternatives – both may jointly and simultaneously contribute to the outcome (“I believe it’s a combination of the two,” P says explicitly), and the dialogue then turns to the relative efficacy of the two, again with an emphasis on empirical data as the basis for a judgement and recognition that data may weaken as well as support a causal claim.

A third characteristic is also evident. Both P and N represent the dialogue at a meta-level – they make repeated reference to what they are doing and seek to accomplish. When N acknowledges “You have a point”, the subject is the dialogue itself and the relation between the speakers’ respective claims, rather than voicing of the claims themselves. P makes an even more ambitious meta-level effort to identify this relation: “I agree the government has a responsibility to stop people. I think we just disagree on the means by which they do this.”

The dialogue shown in Table 6 by contrast shows none of these characteristics. The two speakers alternate turns, each presenting their preferred causal candidates, with gradual elaboration seeking to make their positions more convincing. Absent is any reference to actual or potential evidence that would support the causal claim being advanced. Equally critical, neither member of the pair directly addresses the other’s claims, instead using the conversational turn to elaborate the speaker’s own claim. Only at turn 5 does A first address O’s claim of monetary cause by denying its causal status (“Teachers don’t

work for pay”), with O responding by reasserting its efficacy. This pattern occurs again, with A repeating the same denial (“Teachers do not get into this field because of wages”). Following another such repetition, A offers the first and only counterargument to support rejection of O’s claim: “Teachers go into the profession with a general idea of the salary”, a factor that does not hold for A’s preferred cause.

Nor at any point does either speaker evidence meta-level awareness that both their causes could be operating. Because A and O’s dialogue has minimal transactive structure, or overall structure beyond that of alternating turns, it is not surprising that absent from the dialogue is meta-level discourse about the exchange itself (beyond an unelaborated non-acceptance of the other’s claim: “Not sure if I agree”). It may be the case that A and O see no function of the dialogue beyond that of airing their respective views, which they of course could have done as well individually outside a dialogic context. In these respects, then, their dialogue reflects the failed or at least compromised interchange that may occur when the characteristics observed in P and N’s dialogue are absent.

Although the dialogues in [Tables 5](#) and [6](#) are strikingly different, along the several dimensions we identify, qualitative study of individual cases can of course never be more than suggestive, and we offer such qualitative analysis here only in this vein, for the purpose of illustration. We nonetheless see it as important in identifying such potential dimensions in terms of which discourse of the genre investigated may vary, as candidate dimensions for further examination. In the analysis we turn to next, we take a different, more quantitative approach to establishing a connection between patterns of reasoning and patterns of discourse.

### *Quantitative group data*

In the Study 3 quantitative analysis, we examine the extent to which the pattern participants displayed on our instrument was predictive of their highly discourse dependent performance in the course in which they were enrolled. The course instructor agreed to make available to us each small group’s score on the key performance variable, the stock price index that their group had attained by the end of the simulation. An index of 2000 or higher, the instructor indicated, reflects strong performance, while scores of 1500 or below reflect weak performance. Other measures of a group’s performance exist, but because these were highly correlated with stock price index, only this outcome variable is used as the basis for our analysis.

Of the 25 groups, 11 showed strong performance in the simulation according to the instructor’s criterion, while 7 groups showed intermediate performance and 7 groups showed weak performance. We, therefore, examined the constellation of causal reasoning performance on our measure shown by the group members of each of the groups in these three categories.

In the category of strong group performance, of the 11 groups in this category, 9–82%–contained either (a) at least two members who showed a strong preference for Option C (chose it consistently across all three items), or (b) at least one member who showed a strong preference for Option C (chose it consistently across all three items) and at least three members who showed at least a dominant preference for Option C (chose it for two or more of the three items).

By comparison, of the seven groups whose group performance fell in the intermediate category, three (43%) met this criterion. Of the seven groups whose group performance fell in the weak category, only one (14%) met this criterion. The 11 groups in the top category differed significantly from the 14 groups in lower two categories with respect to proportion who met this criterion,  $\chi^2 = 7.00$ ,  $df = 1$ ,  $p = 0.008$ .

The 11 groups in the top category were more likely to have four members (45% did) than were groups in the intermediate category (where 14% did) or the weak category (where 29% did) – a difference that could be explained by the greater likelihood that a four-person group contained two or three members that met the above criterion (while not ruling out other explanations for superior performance of four-member groups).

There was no difference across the three categories of groups in likelihood of containing one or more members who showed a dominant or strong preference for Option A or for Option B. These percentages were 64%, 57% and 57%, across the high, intermediate and low performing categories, respectively. Containing a member who was a less capable causal reasoner thus appeared to not harm a group's performance, with the important proviso that the group contained at least two individuals who met our criterion for classification as a strong multivariable causal reasoner.

Anecdotal evidence supporting this conclusion is the fact that only one group contained only one strong causal reasoner (who consistently preferred Option C) coupled with two members who preferred other than Option C (both preferred Option B). As part of the final class evaluation questionnaire students were asked if they were satisfied with the group process and students uniformly reported high satisfaction, claiming that they made better decisions as a group than they would have done alone. The only exception was an individual in the group just indicated, containing the strong reasoner coupled with two weaker ones; in this group, the strong reasoner rated the group process as unproductive.

## Discussion

We cannot conclude with certainty that groups who lacked a majority of strong reasoners on our instrument for this and only this reason did not as a group perform as well as groups who had this characteristic. Yet the task



demands these groups faced with respect to multivariable causal reasoning make them especially likely to need and benefit from such reasoning. Our findings suggest that to make the soundest decisions and thereby do well, a group needs not just a strong reasoner who will incorporate the operation of multiple factors potentially affecting an outcome into inferences and consequent decisions. This individual needs to engage with at least one other who thinks in these same terms – that is, coordinating causal factors, rather than dropping one in favour of another, and seeking evidence of their operation rather than being satisfied with simply their plausibility.

## General discussion

Together these studies meet our goal of indicating an empirical, as well as conceptual, link between accomplished multivariable causal reasoning and quality of discourse. Identifying individual reasoning skill as a potential contributing factor to discourse quality is important for multiple reasons, not the least among them because of its implications for ways such failure might be addressed. The discourse process itself, to be sure, warrants attention with regard to remediation, but so, we have suggested here, do the reasoning capabilities and dispositions of the individuals who engage in the process.

We have in work going back several decades observed argumentative reasoning, as well as argumentative discourse, to be wanting in ways consistent with the weaknesses identified in the present investigation (Kuhn, 1991; Felton & Kuhn, 2001; Modrek & Kuhn, 2017). Attention to the other's claims is often missing in favour of focus on one's own claims, and attempts at counterargument often take the form of posing alternatives rather than addressing the claim at hand. Missing in such discourse is what Walton (1989, 2014) identifies as one of two core objectives of argumentation – seeking to weaken the opponent's claims.

If the interlocutors in an argumentative dialogue behave in a comparable way, the two literally argue past one another. The qualitative illustrations we presented in Study 3 are suggestive of a broader point. The partners P and N, we can surmise, have gained in their respective understanding of their topic as a consequence of their discussion. This appears less likely to be true in the case of partners A and O.

On a more positive note, at least in the case of adolescents, research indicates that competence is indeed modifiable. Simply with engagement and dense practice within a peer group over an extended period, young adolescents begin to show more rigorous standards of evidence and of inference in their argumentation (Crowell & Kuhn, 2014), and adults show progress with even more minimal interventions (Zavala & Kuhn, 2017). Moreover, they come to expect adherence to these norms of one another (Kuhn, Zillmer, Crowell, & Zavala, 2013). Discourse with a more capable partner strengthens this

effect (Papathomas & Kuhn, 2017; Mayweg-Paus, Macagno, & Kuhn, 2016). These developments in discourse skill, moreover, transfer to individual written essays on both the discourse topics and new topics (Kuhn & Crowell, 2011; Hemberger, Kuhn, Matos, & Shi, 2017). These positive findings are all the more important, given the indication from Studies 1 and 2 that little development can be expected from early adolescence to adulthood in this regard, in the absence of targeted intervention.

Similarly, with engagement and extended practice in causal inference participants show indications of relinquishing a univariable model of causality in favour of a multivariable one (Kuhn, Ramsey, & Arvidsson, 2015). In particular, we have seen a decrease in types A and B and an increase in type C in high school students' judgements following an extended intervention in which they gained practice in using evidence to support and weaken their claims (Kuhn, Arvidsson, Lesperance, & Corprew, 2017). Whether we would be as successful with older individuals remains to be seen.

In conclusion, it remains to be noted that the claims made here provide no reason to discount the importance of social and cultural trends in what many modern observers point to as a decline in the quality of public discourse. Social climate indeed counts for much in supporting and sustaining intellectual discourse. Yet seeking to promote the intellectual competence and values of individuals, as well as social groups and societies, can only play a positive role.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## References

- Ahn, W., Kalish, C., Medin, D., & Gelman, S. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, 54, 299–352.
- Brem, S., & Rips, L. (2000). Explanation and evidence in informal argument. *Cognitive Science*, 24, 573–604.
- Crowell, A., & Kuhn, D. (2014). Developing dialogic argumentation skills: A three-year intervention study. *Journal of Cognition and Development*, 31, 456–496.
- Felton, M., & Kuhn, D. (2001). The development of argumentive discourse skill. *Discourse Processes*, 32, 135–153.
- Goedert, K., Harsch, J., & Spellman, B. (2005). Discounting and conditionalization: Dissociable cognitive processes in human causal inference. *Psychological Science*, 16, 590–595.
- Goksun, T., George, N., Hirsh-Pasek, K., & Golinkoff, R. (2013). Forces and motion: How young children understand causal events. *Child Development*, 84, 1285–1295.
- Hemberger, L., Kuhn, D., Matos, F., & Shi, Y. (2017). A dialogic path to evidence-based argumentive writing. *Journal of the Learning Sciences*, 26, 575–607. doi:10.1080/10508406.2017.1336714

- Jervis, R. (2006). Understanding beliefs. *Political Psychology*, 27, 641–663.
- Kelley, H. (1973). The processes of causal attribution. *American Psychologist*, 28, 107–128.
- Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge: MIT Press.
- Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.
- Kuhn, D. (2001). How do people know? *Psychological Science*, 12, 1–8.
- Kuhn, D. (2009). The importance of learning about knowing: Creating a foundation for development of intellectual values. *Perspectives on Child Development*, 3, 112–117.
- Kuhn, D., Arvidsson, T. S., Lesperance, R., & Corprew, R. (2017). Can engaging in science practices promote deep understanding of them? *Science Education*, 101, 232–250.
- Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science*, 22, 545–552.
- Kuhn, D., & Katz, J. (2009). Are self-explanations always beneficial? *Journal of Experimental Child Psychology*, 103, 386–394.
- Kuhn, D., Ramsey, S., & Arvidsson, T. S. (2015). Developing multivariable thinkers. *Cognitive Development*, 35, 92–110.
- Kuhn, D., Zillmer, N., Crowell, A., & Zavala, J. (2013). Developing norms of argumentation: Metacognitive, epistemological, and social dimensions of developing argumentative competence. *Cognition & Instruction*, 31, 456–496.
- Laux, J., Goedert, K., & Markman, A. (2010). Causal discounting in the presence of a stronger cue is due to bias. *Psychonomic Bulletin & Review*, 17, 213–218.
- Mayweg-Paus, E., Macagno, F., & Kuhn, D. (2016). Developing argumentation strategies in electronic dialogs: Is modeling effective? *Discourse Processes*, 53, 280–297.
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34, 57–111.
- Modrek, A., & Kuhn, D. (2017). A cognitive cost of the need to achieve? *Cognitive Development*, 44, 12–20.
- Moshman, D. (2015). *Epistemic cognition and development: The psychology of justification and truth*. New York, NY: Psychology Press.
- Papathomas, L., & Kuhn, D. (2017). Learning to argue via apprenticeship. *Journal of Experimental Child Psychology*. Advance online publication. doi:10.1016/j.jecp.2017.01.013
- Ricco, R. (2015). The development of reasoning. In L. Liben & U. Mueller, (Eds.), R. Lerner (Series Ed.), *Handbook of child psychology and developmental science* (Vol. 2, 7th ed.). Hoboken, NJ: Wiley.
- Richter, T., & Schmid, S. (2010). Epistemological beliefs and epistemic strategies in self-regulated learning. *Metacognition and Learning*, 5, 47–65.
- Rips, L. (1998). Reasoning and conversation. *Psychological Review*, 105, 411–441.
- Rittle-Johnson, B., & Loehr, A. (2016). Eliciting explanations: Constraints on when explanation aids learning. *Psychonomic Bulletin & Review*. Advance online publication. doi:10.3758/s13423-016-1079-5
- Schustack, M., & Sternberg, R. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General*, 110, 101–120.
- Slooman, S. (2005). *Causal models: How people think about the world and its alternatives*. New York, NY: Oxford University Press.
- Stanovich, K. (2009). *Decision making and rationality in the modern world*. Oxford: Oxford University Press.
- Stanovich, K. (2010). *What intelligence tests miss: The psychology of rational thought*. New Haven: Yale University Press.

Tetlock, P., Vieider, F., Oatil, S., & Grant, A. (2013). Accountability and ideology: When left looks right and right looks left. *Organizational Behavior and Human Decision Processes*, 122, 22–35.

Van Eemeren, F. H., Garssen, B., & Meuffels, B. (2009). *Fallacies and judgments of reasonableness. Empirical research concerning the pragma-dialectical discussion rules*. Dordrecht: Springer.

Walton, D. (1989). Dialogue theory for critical thinking. *Argumentation*, 3, 169–184.

Walton, D. (2014). *Dialog theory for critical argumentation*. Amsterdam: John Benjamins.

Williams, J., Lombrozo, T., & Rehder, R. (2013). The hazards of explanation: Overgeneralization in the face of exceptions. *Journal of Experimental Psychology*, 142, 1006–1014.

Zavala, J., & Kuhn, D. (2017). Solitary discourse is a productive activity. *Psychological Science*, 28, 578–586. doi:10.1177/0956797616689248

### Appendix. Item analysis of Study 2 data

1. Some health officials have found cancer rates higher in cities than in outer areas. Dr J. Rawls claimed tanning salons are to blame. Choose the best statement to use if you wanted to argue this doctor was wrong.

Response options	Adolescents		Adults	
A. Air pollution is a more likely cause of cancer in the city.	n = 21	39%	n = 8	22%
B. Many people who don't go to tanning salons also get cancer	n = 25	46%	n = 23	64%
C. Many people outside the city also go to tanning salons and don't get cancer.	n = 8	15%	n = 5	14%

2. People from some countries have longer average life expectancy than people in others. Dr F. Cole claimed a diet high in fish causes long life. Choose the best statement to use if you wanted to argue this doctor was wrong.

Response options	Adolescents		Adults	
A. Exercise is a more important cause of long life.	n = 32	59%	n = 16	44%
B. People who don't eat fish often live to an old age.	n = 14	26%	n = 17	47%
C. People who eat a lot of fish often live only to an average age.	n = 8	15%	n = 3	9%

3. Venezuela is a country with money trouble, unable to pay its bills. Dr P. Garet claimed the cause was too many social programs to help people. Choose the best statement to use if you wanted to argue this doctor was wrong.

Response options	Adolescents		Adults	
A. Poor money management is a more likely cause of a country's money trouble.	n = 14	26%	n = 15	42%
B. Some countries like Haiti have very few programs to help their people and Haiti has serious money shortages.	n = 18	33%	n = 7	19%
C. Some countries like Sweden have many social programs and are not in money trouble.	n = 22	41%	n = 14	39%

4. Diabetes is an increasing health problem. A doctor has claimed that lack of exercise is a cause. Choose the best statement to use if you wanted to argue this doctor was wrong.

Response options	Adolescents		Adults	
A. Poor diet is a more likely cause of diabetes.	<i>n</i> = 28	52%	<i>n</i> = 15	42%
B. Some people who exercise daily develop diabetes.	<i>n</i> = 12	22%	<i>n</i> = 11	30%
C. Some people never exercise and don't develop diabetes.	<i>n</i> = 14	26%	<i>n</i> = 10	28%

5. Many big cities have high teen crime rates. A politician has claimed that lack of social programs for teens is a cause. Choose the best statement to use if you wanted to argue this politician was wrong.

Response options	Adolescents		Adults	
A. Poor schools are a more likely cause of teen crime.	<i>n</i> = 13	24%	<i>n</i> = 6	17%
B. Some cities have social programs for teens and have high teen crime.	<i>n</i> = 26	48%	<i>n</i> = 20	56%
C. Some cities have no social programs for teens but have little teen crime.	<i>n</i> = 15	28%	<i>n</i> = 10	27%

6. Some states have high murder rates. A politician has claimed that lack of the death penalty in those states is the cause. Choose the best statement to use if you wanted to argue this politician was wrong.

Response options	Adolescents		Adults	
A. A life jail sentence is a more effective punishment for murder.	<i>n</i> = 27	50%	<i>n</i> = 8	22%
B. Some states that have the death penalty have high crime rates.	<i>n</i> = 15	28%	<i>n</i> = 18	50%
C. Some states that don't allow the death penalty have low crime rates.	<i>n</i> = 12	22%	<i>n</i> = 10	28%

7. Some schools have very strict behaviour rules. Leaders of these schools claim strict rules are the cause of children's high achievement. Choose the best statement to use if you wanted to argue this was wrong.

Response options	Adolescents		Adults	
A. Teachers are a more important factor in students' achievement.	<i>n</i> = 15	28%	<i>n</i> = 14	39%
B. Some schools have few rules and their students achieve well.	<i>n</i> = 19	35%	<i>n</i> = 7	19%
C. Some schools have strict behaviour rules and their students do poorly.	<i>n</i> = 20	37%	<i>n</i> = 15	42%

8. A head of a social work agency claims that growing up in poverty causes psychological problems in adulthood. Choose the best statement to use if you wanted to argue this was wrong.

Response options	Adolescents		Adults	
A. Growing up in a single parent home is a more important cause of psychological problems in adulthood.	<i>n</i> = 17	31%	<i>n</i> = 9	25%
B. Children from middle and high-income families often have psychological problems in adulthood.	<i>n</i> = 14	26%	<i>n</i> = 18	50%
C. Children who grow up in poverty often become well-adjusted adults.	<i>n</i> = 23	43%	<i>n</i> = 9	25%

Note: Items 1–3 are used in Studies 1 and 3; items 1–8 in Study 2.