

# Arguments about Deletion

## How Experience Improves the Acceptability of Arguments in Ad-hoc Online Task Groups

**Jodi Schneider**

Digital Enterprise Research  
Institute, National University  
of Ireland, Galway  
jodi.schneider@deri.org

**Krystian Samp**

Digital Enterprise Research  
Institute, National University  
of Ireland, Galway  
krystian.samp@deri.org

**Alexandre Passant**

Digital Enterprise Research  
Institute, National University  
of Ireland, Galway  
alexandre.passant@deri.org

**Stefan Decker**

Digital Enterprise Research  
Institute, National University  
of Ireland, Galway  
stefan.decker@deri.org

### ABSTRACT

Increasingly, ad-hoc online task groups must make decisions about jointly created artifacts such as open source software and Wikipedia articles. Time-consuming and laborious attention to textual discussions is needed to make such decisions, for which computer support would be beneficial. Yet there has been little study of the argumentation patterns that distributed ad-hoc online task groups use in evaluation and decision-making. In a corpus of English Wikipedia deletion discussions, we investigate the argumentation schemes used, the role of the arguer's experience, and which arguments are acceptable to the audience. We report three main results: First, the most prevalent patterns are the *Rules* and *Evidence* schemes from Walton's catalog of argumentation schemes [34], which comprise 36% of arguments. Second, we find that familiarity with community norms correlates with the novices' ability to craft persuasive arguments. Third, acceptable arguments use community-appropriate rhetoric that demonstrate knowledge of policies and community values while problematic arguments are based on personal preference and inappropriate analogy to other cases.

### Author Keywords

collaboration and conflict; argumentation schemes; critical questions; online argumentation; decision-making; deliberation; peer production; Wikipedia

### ACM Classification Keywords

H.5.3 Information Interfaces and Presentation: Group and Organizational Interfaces—*Collaborative computing*

### INTRODUCTION

Jointly creating digital artifacts in a distributed collaboration is commonplace, for instance in distributed software devel-

opment [9] and collaborative authoring [10]. Distributed online communities which produce jointly-created artifacts often need to make decisions on these artifacts, such as when an open source software product is ready for release, when a W3C recommendation is completed, or whether a Wikipedia article is acceptable for continued inclusion in the encyclopedia. Collaborators may at first rely on task independence, formal coordination, or shared mental models (sometimes acquired through informal discussion). After the creation phase, artifacts may move into revision and acceptance/rejection phases, which may draw discussion from larger groups.

Decision-making often relies on generating and integrating feedback from a large group, particularly for the products of diverse, widely distributed online communities. However, these feedback discussions can be overwhelming, making it difficult to focus the discussion on the relevant aspects. Further, each community applies specific criteria to determine when artifacts are *ready to release* and *appropriate*, and what kinds of comments are relevant and helpful at a given point, meaning that only certain kinds of arguments are relevant and acceptable. Different viewpoints, from multiple people, may be submitted for consideration, and when disagreements arise, in order to make progress, resolutions must be found. The process of coming to agreement should:

- consider all relevant arguments
- be inclusive of multiple viewpoints
- minimize effort
- avoid alienating participants
- result in a high quality decision.

The English-language Wikipedia is a prime example of a collaborative authoring community, where ad-hoc voluntary groups undertake a number of tasks such as creation, editing, and deletion [4]. Deletion is a prominent activity involving significant conversations within and at the boundaries of the community [32, 20], which is made more challenging since newcomers often misunderstand the policies, such as

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CSCW'13, February 23–27, 2013, San Antonio, Texas, USA.

Copyright 2012 ACM 978-1-4503-1209-7/13/02...\$15.00.

Wikipedia’s notability<sup>1</sup> policy<sup>2</sup>. Despite numerous guidelines and a standardized workflow, disagreements are evident in deletion discussions [5]. There are an average of 500 discussions a week, and some articles are repeatedly discussed [32]. Therefore, deletion discussions are a useful first place to analyze the disagreement resolution process in order to identify common patterns.

Our work has three primary goals:

1. to identify which arguments are given in content deletion discussions
2. to investigate the differences between novices’ and experts’ arguments
3. to describe the structure of acceptable arguments both for keeping and for deleting content.

The outcomes of our work can be applied in several ways. Argumentation patterns, such as the argumentation schemes used in this paper [34], are needed to apply existing work on decision-making and group support systems (e.g. [11]). For socializing newcomers, externalizing acceptable argumentation patterns and making them explicit may be particularly helpful; as we show, acceptance of arguments depends in part on how well they display an understanding of community policies and values. Further, we provide evidence about the role of argumentation in online decision-making, an important topic with little previous investigation.

Following this introduction, we first present related work. Subsequently we describe the deletion process and then present our research questions and methodology. We then analyze typical arguments, explain how experience affects arguments, and describe what types of arguments are accepted. Finally, we conclude with discussion and future work.

## RELATED WORK

We now review related work about team behavior, online decision-making, deletion in Wikipedia, and newcomers.

### Team behavior

Team behavior has been studied in various research domains, including psychology, management science, and CSCW. Kozlowski & Ilgen review the psychology literature on groups and teams, which indicates that some types and levels of conflict are helpful in groups; that team processes and task demands should be appropriately matched; and that shared and team mental models may be important in effective task completion [22]. Further, turn-taking is important in effective group performance, at least for synchronous, in-person groups, and social sensitivity impacts the group productivity [35]. Research on virtual teams has found that “routine patterns of media use” are quickly established and thereafter are maintained, which provides a first mover advantage to establishing communication procedures [19]. Our work addresses the routine procedures used in arguing, and requires an understanding of what makes task completion successful.

<sup>1</sup><http://en.wikipedia.org/wiki/WP:Notability>

<sup>2</sup>Notability issues account for over one-fifth of all deletions from the encyclopedia, as articles that do not articulate their importance are summarily deleted in great numbers without discussion or controversy [25, 15].

## Online decision-making

Numerous groups – notably open source communities and some standardization bodies (e.g. the W3C<sup>3</sup>) – use online discussions as their primary medium for decision-making. The rhetoric of decision-making has already received some attention in these communities. For instance Chilana, Ko, and Wobbrock investigated rhetorical devices and contentiousness in bug reports, in order to raise design criteria for future bug reporting [7]. In the Python community, Barcellona, Détienne, and Burkhardt have investigated how design proposals are moved forward to successful implementation through boundary-spanning participants who are able to bridge the rhetoric of novices and experts [2]. In comparison to the typical speech-act oriented approaches, the process for argumentative analysis we describe in this paper is intensive and adds detail; in the future it could be applied to any epistemic or artifact-oriented community.

## Previous Research on Deletion in Wikipedia

The arguments used in deletion discussions are of primary importance: according to community policy, decisions should be made by consensus, based on which arguments prevail. However, previous research on deletion in Wikipedia has focused primarily on the votes given and the numbers supporting each side of a discussion. Taraborelli and Ciampaglia investigated the chronological order of votes in controversial deletion discussions [32]. Schneider et al. investigated decision factors in a representative sample of deletion discussions [30]. Geiger [15] and Lam et al. [25] documented deletion statistics based on keyword logs. The bulk of decisions are made by the same regular participants: Article creators rarely (18%) participate [15]. Only 26% of all deletion debates include newcomers, and only 8% include more than one newcomer [15]. However, researchers have suggested that ensuring broad participation may be important for fighting discussion bias, one of the shortcomings of the deletion process [32]. Decisions made by large groups with diverse tenure are least likely to be overturned, while decisions made with the participation of newcomers are more likely to be reversed [24]. Our work contributes by focusing on the arguments used to make decisions and by identifying problematic aspects of newcomers’ participation in deletion discussions.

## Newcomers and Socialization

We now review work on newcomers in online communities, and in particular, in Wikipedia. Five problems of dealing with newcomers are addressed in [23]: recruitment, selection, retention, protection, and socialization. Newcomer socialization, and the differences between newcomer and expert performance, have been a frequent topic of research. Various socialization techniques have been tested. For instance, goal-setting, time-tables, and feedback (both personalized emails and anonymous peer feedback), along with recommended worked examples, were used to socialize newcomers to a tax Q&A community in [12].

In Wikipedia, socialization is of heightened importance because of Wikipedia’s slowing growth [31, 17]. Yet socializa-

<sup>3</sup>See i.e. the HTML standards group Decision Policy: <http://dev.w3.org/html5/decision-policy/decision-policy-v3.html>

tion may also be at odds with novices' perception of the encyclopedia as an artifact, rather than a community [4]. Various of Wikipedia's socializing mechanisms have been studied, including legitimate peripheral participation [1], group membership [8, 13], and inter-editor communication [27], which can function as a deterrent to further participation as well as a socialization tactic [16]. In general, as Kraut et al. have noted, there is a tension between protecting the community (since newcomers do not know and may not be motivated to conform to community norms) and welcoming and socializing newcomers [23]. This is particularly evident when deleting content: participants are more likely to leave the encyclopedia project when content they added is deleted.<sup>4</sup> In quantifying differences in novices' arguments, the present study points towards interventions; we also highlight the need to maintain quality without alienating participants.

## THE DELETION PROCESS

Next, we describe the deletion process, the focus of our work. Anyone can suggest that a Wikipedia article be deleted, by editing it and adding a special flag. Three main deletion procedures are used, depending on how controversial deletion is expected to be and the impact of deleting material. Under *speedy deletion*, no waiting period is required before an administrator deletes clearly inappropriate content (e.g. vandalism and spam). *Proposed deletion* is meant for uncontroversial cases; the deletion notice must remain uncontested for a seven day waiting period. *Articles for Deletion*<sup>5</sup> is intended for controversial articles.

### Articles for Deletion Discussions

Articles for Deletion (AfD) is the most deliberative of Wikipedia's three main deletion procedures, involving community discussions to determine whether controversial articles are appropriate topics for the encyclopedia. A sample AfD discussion is shown in Figure 1. In each AfD discussion, a nominator gives a justification for deleting an article; the community discusses the merits of the article and topic, providing arguments for or against deleting the article; and a discussion closer—generally an administrator—reviews the discussion after seven days, with the intention of finding a consensus decision.

Our work focuses on AfD since these community discussions have the longest and most elaborate argumentative discussions and thus enable us to best study the deliberation process. Nominators, discussants, and closers are self-selected. While these deletion discussions are open to anyone – even IP users without a username – to read and to comment on, they are sophisticated wiki spaces with their own conventions: messages start with a bolded indication of their 'vote'<sup>6</sup> (*Keep*, *Delete*, *Merge*, etc.), they are signed with the poster's username or IP

<sup>4</sup>“only 0.6 percent of those whose articles are met with deletion stayed editing, compared to 4.4 percent of the users whose articles remained” [http://en.wikipedia.org/w/index.php?title=Wikipedia:Wikipedia\\_Signpost/2011-04-04/Editor\\_retention&oldid=425153383](http://en.wikipedia.org/w/index.php?title=Wikipedia:Wikipedia_Signpost/2011-04-04/Editor_retention&oldid=425153383)

<sup>5</sup>[http://en.wikipedia.org/wiki/WP:Articles\\_for\\_deletion/](http://en.wikipedia.org/wiki/WP:Articles_for_deletion/)

<sup>6</sup>Colloquially referred to as *!vote*, to emphasize that they are *not* votes.

**Heath Totten** [edit][vote]

Heath Totten (edit | talk | history | links | watch | logs) – (View log)  
 (Find sources: "Heath Totten" – news · books · scholar · JSTOR · free images)

**Nomination** Alex (talk) 22:13, 29 January 2011 (UTC)  
 Non-notable, 32-year old, free agent, minor league baseball pitcher who hasn't played since 2008. His 66-73 record is far from stellar and, in my opinion, does not merit an article.

**Message** Keep (talk) 22:38, 29 January 2011 (UTC)  
 "hasn't played since 2008" His page at the official Minor League Baseball website [1] states that his status is "Active". It also states that he has pitched as recently as December 29, 2010. He is playing for Bravos de Margarita in the Venezuelan League - the highest professional league in the country of Venezuela. Some of his teammates this past year are major league players. The second guideline of the baseball notability guidelines: "Have appeared in at least one game in ... any other top-level national league." Having played in the top professional league of Venezuela, I feel he qualifies.

**Discussion** Reply Spanneracl (talk) 22:49, 29 January 2011 (UTC)  
 I would also note that he participated in the 2010 Caribbean Series. Wikipedia:Notability (sports) states that "Sports figures are presumed notable (except as noted within a specific section) if they: 1. have participated in a major international ... professional competition at the highest level"

The Olympics are "major." The Caribbean Series is not. Alex (talk) 18:38, 31 January 2011 (UTC)

The Olympics are a major amateur competition. The Caribbean Series is professionals only. Regardless, there were no Summer Olympics in 2010, so one could argue that this was the most important international baseball tournament of that year - professional or amateur. Kinston eagle (talk) 19:31, 31 January 2011 (UTC)

Keep per Kinston. Spanneracl (talk) 23:07, 29 January 2011 (UTC)

'Vote' Rationale Username Timestamp

Signature

Figure 1. An extract from the deletion discussion for baseball player Heath Totten. Community discussion follows the nomination; participants can reply to the nomination or to each other. A typical message consists of a 'vote' followed by a rationale, signed with the poster's username and a timestamp.

address, and, most importantly, messages must use appropriate rhetoric based on Wikipedia guidelines and policies.

While these 'votes' may be helpful, majority vote does not in fact determine the outcome. Rather, in the English-language Wikipedia, decisions are intended to be made by consensus, based on whether one side "substantively defeated the other's key arguments"<sup>7</sup>. Yet these actual arguments made in deletion discussions have not been investigated.

## RESEARCH QUESTIONS

We consider three research questions about content deletion. In a corpus of user discussions called Articles for Deletion, used to decide what topics deserve articles in the English Wikipedia, we ask:

- [RQ1] What arguments are given?
- [RQ2] Do people with different levels of experience with Wikipedia editing or the Wikipedia deletion process provide different types of arguments?
- [RQ3] Which argumentation schemes are accepted?

Next we describe the argument classification used in this study.

<sup>7</sup>Policy from [http://en.wikipedia.org/wiki/Wikipedia:Deletion\\_process#Consensus](http://en.wikipedia.org/wiki/Wikipedia:Deletion_process#Consensus), well-summarized by an administrator's closing remarks at [http://en.wikipedia.org/WP:Articles\\_for\\_deletion/List\\_of\\_common\\_misconceptions.\(3rd\\_nomination\)](http://en.wikipedia.org/WP:Articles_for_deletion/List_of_common_misconceptions.(3rd_nomination))

## ARGUMENTATION SCHEMES

In examining which arguments are given, we followed philosopher Douglas Walton's 2008 classification [34]. With 60 argumentation schemes collected from other scholars and from his own work, Walton's collection is the most comprehensive. Refined over more than a decade, Walton's schemes are widely used in computational argumentation since they "have provided exactly the right balance between theoretical consistency... and practical utility" [29]. Walton traces the modern use of argumentation schemes to 1953 and his own first collection of 25 argumentation schemes was published in 1995 [33], which originated in studying under what conditions traditional logical fallacies were in fact "appropriate, acceptable, and persuasive" (p11) [34].

### Example Argumentation Scheme

Argumentation schemes classify the patterns used in making arguments. Walton's schemes each begin with a title and a description. For example, here is the argumentation scheme **Argument from Rules - From Established Rule** [34]:

*Major Premise:* If carrying out types of actions including A is the established rule for x, then (unless the case is an exception), a must carry out A.

*Minor Premise:* Carrying out types of actions including A is the established rule for a.

*Conclusion:* Therefore, a must carry out A.

To indicate possible flaws in reasoning, associated with each scheme, there are critical questions pointing to the possible counterarguments. For example, three *critical questions* accompany the Argument from Rules:

1. Does the rule require carrying out this type of action?
2. Are there other established rules that might conflict with or override this one?
3. Are there extenuating circumstances or an excuse for non-compliance?

Such critical questions can be used to guide a debate, and to ensure that all possible objections have been addressed.

### Instantiating an Argumentation Scheme

Argumentation schemes are *patterns* for arguing: to use them, they must be instantiated with details. We now provide a concrete example instantiating the above example argumentation scheme Argument from Rules. We draw a (hopefully familiar) example from the rules of the road: the rule that vehicles must stop at red lights.

If stopping at a red light is the established rule for driving a vehicle, then (unless the case is an exception), drivers must stop at a red light.

Stopping at a red light is the established rule for drivers.  
Therefore, drivers must stop at a red light.

Using this example, we also provide some instantiations of the critical questions above:

1. Were you driving a vehicle?
2. Did a police officer direct you to continue without stopping?
3. Were you driving an ambulance with its siren on?

Next we examine argumentation schemes in context.

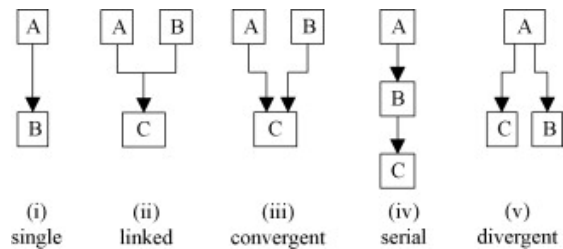


Figure 2. To reason from premises to conclusions, argumentation schemes (represented by arrows) are used. An argument may involve one or several premises, conclusions, and argumentation schemes. Diagram from Rahwan [28].

## EXAMPLE ARGUMENT

Figure 1 shows the beginning of a debate about baseball player Heath Totten<sup>8</sup>. Six messages are shown: the nomination (1 message), a bolded 'Keep' vote (1 message), with three replies indented below it (3 messages), and a second bolded 'Keep' vote (1 message). We use this as an example for analysis below.

### Analysis

This argument for keeping the article is fully articulated. It is principally an application of the Argument from Rules described above, as a policy interpretation about which guideline should be applied (baseball notability<sup>9</sup>). It combines this with an Argument from Evidence (see Table 1)—a sourced, factual correction (the player is currently active). Additional comments (e.g. *Some of his teammates this past year are major league players*) are given. Unusually, in this case, the connection between the rule and the evidence are made explicit: *Having played in the top professional league in Venezuela, I feel he qualifies.*

A second follow-up response, from the same person, is more typical, in that it is less explicit. It again uses the same two argumentation schemes—a Rule (baseball notability) and Evidence (played in the Caribbean series)—yet without spelling out the further claim that this evidence shows that the player meets the baseball notability guideline. As typical in conversations, this is an enthymeme: the arguer does not explicate the argument fully, and the reader must infer the implied claim that the player meets the sports guideline. Unstated and missing information adds challenges for machine processing. Yet, as evidenced by the nominator's response, for a human, there is no ambiguity here. Thus, the nominator replies, contesting that the Caribbean series qualifies under the baseball notability guideline.

In general, multiple arguments may be given in a message, and arguments can be used in various combinations, drawn from Walton's schemes (Table 1). Arguments can be related as supporting sub-arguments, or can be independent, separate arguments, as shown in Figure 2 (from [28]).

## METHODOLOGY

We used Walton's 60 argumentation schemes (see the example above) as the basis for a categorization of the argu-

<sup>8</sup>[http://en.wikipedia.org/wiki/Wikipedia:Articles\\_for\\_deletion/Heath\\_Totten](http://en.wikipedia.org/wiki/Wikipedia:Articles_for_deletion/Heath_Totten)

<sup>9</sup>[http://en.wikipedia.org/wiki/Wikipedia:Notability\\_\(sports\)#Baseball](http://en.wikipedia.org/wiki/Wikipedia:Notability_(sports)#Baseball)

ments appearing in a sample corpus of deletion. To address our research questions, after determining the argumentation classifications to use, we needed a representative sample. Within this sample we needed to count the arguments presented (RQ1), determine which users were novices and what arguments they gave (RQ2), and determine which arguments were challenged (RQ3). Choosing the level at which to annotate arguments was particularly challenging, since messages can contain any number of arguments, or no argument at all. For this paper, we annotated arguments at two different levels: first, to determine which arguments were most prevalent, we used all arguments, constituted of a clause, several sentences, or the entire message. Later, to determine the main argument, we coded one main argument per message. We now detail our methodology.

### Representative Sample

Our core corpus was the 72 debates begun or relisted on January 29, 2011<sup>10</sup>; this is a typical day, with an average number of debates<sup>11</sup>. Figure 1 shows a nomination and responses from a debate in our corpus. Debates include from 2 to 33 messages (stddev 6). This core corpus consists of 741 messages contributed by 244 users between the first nomination<sup>12</sup> on January 14 and the last close of debate on February 8. Messages in our core corpus range from about a dozen to over 4,000 characters. The scope of the sample enabled extensive manual examination, unlike previous research which merely reported deletion data based on keyword logs, using Wikipedia's classifications [15] or using rough categories, such as no content/context, notability/significance, and whether the issue was discussed or administratively deleted without discussion [25]. By contrast, we use annotation and content analysis to analyze deletion discussions and outcomes based on hand analysis of our sample, to illuminate the activity of deletion discussions, rather than to provide a fully representative record.

### Iterative Annotation

We annotated arguments in four rounds, following an iterative annotation strategy [18], with the first two rounds of coarse annotation by the principal investigator, and two rounds of subsequent annotation by trained student annotators. Initially, we focused on identifying the most prevalent arguments, in the first and second rounds. To determine the most prevalent argumentation schemes, each message in the corpus was first coded by the principal investigator as a sequence of contiguous arguments selected from Walton's 60 argumentation schemes. Subsequently, the most prevalent schemes were used in the third and fourth rounds; these schemes, along with selections from the final annotation manual are shown in Table 1. To determine the main argument in each message, the same two student annotators selected one of 17 categories characterizing the message's argument strategy. This helped simplify annotation, which aided the crispness of the categorization for the third and fourth rounds, since inter-annotator agreement is particularly sensitive to boundary variation.

<sup>10</sup>[http://en.wikipedia.org/wiki/WP:Articles\\_for\\_deletion/Log/2011\\_January\\_29](http://en.wikipedia.org/wiki/WP:Articles_for_deletion/Log/2011_January_29)

<sup>11</sup>based on January 2007-April 2011

<sup>12</sup>While debates typically last 7 days, they can be relisted for further discussion.

Our annotation guide evolved through this process. For the first round of annotation, 'A User's Compendium of Schemes' (Chapter 9 of [34]) and notes on Walton's 60 schemes were used to annotate a small sample in Corpus-Tool<sup>13</sup>. Based on the data, for the second round of annotation, we added two more categories—*note* and *no reason given*—and prepared our own annotation guide with Wikipedia-based examples, resulting in revisions to the guide and recoding of the sample. Overall argument prevalence (RQ1) was estimated from the second round of annotation, and to scope the further annotation, we revised the annotation guide to comprise 17 categories as shown in Table 1. First, we selected the subset of 14 argumentation schemes that appeared more than 2% of the time. Besides these argumentation schemes, we had three additional categories: *Note* was used for standardized, templated notes used for routine notices. *No reason given* was used for messages that indicate a position without stating an argument. *No argument* was used for messages that do not state an obvious argument. Examples of all 17 classifications are given in Table 1.

Restricting to the most prevalent argumentation schemes, we tested annotation consistency and the annotation guide, in the third round, which used two independent coders. Annotators were an upper-level undergraduate and a graduate student familiar with the corpus from previous annotation tasks with a different categorization on the same corpus [30]. Before the third round of coding, annotators had an initial training meeting with the principal investigator to discuss the annotation guide and tools; during the meeting we independently coded a test sample from a sample corpus and then compared codings. Following discussion and collaborative coding, in the third round, we tested the clarity of the guide by having the two assistants spend 2-3 hours independently coding sample debates from a second, sample corpus<sup>14</sup>, in order to surface problems with the annotation guidelines and examples of hard-to-categorize messages. Subsequently, we met to discuss the annotation scheme and suggest refinements. The principal investigator reviewed the annotated test corpus, examining the argumentation schemes within each category, and listing the messages with discrepancies between the two annotators to provide them with further feedback.

For the fourth round of annotation, an updated manual was prepared, providing further guidance about the categories, based on feedback on the annotation and discriminating examples. Then the same two assistants annotated the original corpus with the final annotation manual, spending about 10 hours on this task. Examples from the manual are shown in Table 1. Interannotator agreement was strong, considering the extreme difficulty inherent in the task, due to the number of categories (17) and the complex analysis involved in determining and categorizing the main argument. Overall, interannotator agreement statistics are shown in Table 2.

To prepare text files for annotation, each debate was downloaded as a single HTML file, then HTML was stripped out

<sup>13</sup><http://www.wagsoft.com/CorpusTool/>

<sup>14</sup>Drawn from debates discussed on March 1, 2012  
[http://en.wikipedia.org/wiki/Wikipedia:Articles\\_for\\_deletion/Log/2012.March\\_1](http://en.wikipedia.org/wiki/Wikipedia:Articles_for_deletion/Log/2012.March_1)

Issues (RQ3)	Argumentation Scheme	Definition	Rough example from Wikipedia	Example from 2012-03-01 corpus
⚠	Argument from Analogy	Based on a similar case.	Articles like X are not notable enough for their own standalone articles.	The test is not for just the combination of those two items, but for the notability of it as a topic. For example, if there are articles on "cars", and "trailers" that doesn't necessarily mean that there should be an article on "cars with trailers" even if that combination of words is common in a google search. That's the question here. I don't know the answer.
	Argument from Bias	Bias suspected.	Reads like an advertisement.	Delete as nothing more than promotional guff by a clearly problematic editor who seems hell-bent on spamming his business and products across Wikipedia.
⚠	Argument from Cause to Effect	Using cause and effect.	Given what's here, it's reasonable to assume that other sources mention X.	Given the vast amount of coverage Apple gets for its product launches I'm sure they will be covered in business courses in the years to come as a good way to do business and get large amounts of attention towards your products.
	Argument from Composition	From the parts to the whole.	Mostly about already discussed at X.	The article is largely surplus to requirements given the information already contained in Battle of Stalingrad and the associated Axis order of battle at the Battle of Stalingrad and Red Army order of battle at the Battle of Stalingrad .
	Argument from Evidence to Hypothesis	Providing evidence.	Source X isn't good enough because... // Here are the sources, which tell us...	Weak keep. Here is some coverage in Publishers Weekly that backs up the "500,000 copies in print" assertion[35]. A writeup in School Library Journal[36]. And some newspaper reviews[37][38][39][40]: the reviewers aren't Michiko Kakutani, but taken together I'd be inclined to keep--preferably as one consolidated article for all the books.
⚠	Argument from Ignorance	Assumption when no supporting evidence can be found.	No search results.	A seemingly unnotable website and blog. It has no references that would support any sort of notability. Searching around only gives results of personal pages (facebook, twitter, etc), thus it fails WP:RS .
	Argument from Need for Help	Help should be provided when possible.	If the article can be fixed through normal editing, then it is not a good candidate for AfD.	Keep - this article needs extensive development and lots of citations, but it is unquestionably notable. It does need some expert attention to select and paraphrase good review articles (i.e. secondary sources) from the thousands of papers on immunity and inflammatory diseases.
	Argument from Position to Know	Personal knowledge.	I grew up in that area & had never heard of her.	Keep: Like Clarityfiend, I know zilch about modern art. But I've actually heard of this guy.
	Argument from Precedent	Based on past decisions.	We've had this same debate for numerous articles, and decided...	Comment by previous closing admin: The deletion requests for this article reminds me of Wikipedia:Articles for deletion/Corn soup. LED-embedded glass is, like corn soup but obviously to a much lesser extent, something almost inherent to any modern metropolitan resident's daily life, hence there is likely to be lots of Google hits but not many of them useful as encyclopedic citations. That said, Google Books did yield some useful results: Popular science magazine, 1986 Structural glass textbook, 2011 Building materials textbook, 2010 and Gizmodo how-to guide As the closing admin of last month's AfD, I don't think it's appropriate for me to !vote here, but as an engineer myself I just want to flag up a few things that may be relevant to this discussion that aren't discussed on the article or the previous AfD.
	Argument from Rules	Rule-based argument.	Fails <guideline>.	Delete . Fails WP:FOOTYN and WP:GNG .
⚠	Argumentation from Values	Evaluate with value judgments.	It's a useful search term, so make it a redirect.	Delete per nom. A list of tallest buildings for a place without any especially tall buildings is pointless and even kind of insulting.
	Argument from Verbal Classification	Definitional arguments.	The current title is misleading.	The concept of immune-mediated inflammatory diseases is not widely recognised within the medical community. Despite what the article says, it is very difficult to group together the widely divergent diseases listed. The fact that they're all treated with immunosuppression is about the only thing. Searching the term as a text word on Pubmed yields 45 references, none of which address the concept in itself. I think deletion is the best step here.
⚠	Argument from Waste	Avoid wasted work.	Merge to save the work. // Delete to save time	This article has been created and then deleted at least three times in the last 12 months. On each occasion it has appeared in a similar format and without much difficulty established that it was self-promotional. On this occasion the author has admitted working for the subject corporation from the get-go. At least they're being honest but this the reliable, independent, standard we aspire too. The article needs to go.
⚠	Practical Reasoning	Actions towards a goal.		Merge with Ledglass, noting that the latter article says "may also be described as LED Glass or LED embedded Glass", and also that the single reference given for this (LED-embedded glass) article doesn't seem to use that phrase but only "Light-emitting diode (LED) illuminated glass".
	No reason given	Vote without explanation.	Per nominator.	Delete per above.
	Note	Non-argumentative note; only templated messages	Note: This debate has been included in the list...	Note: This debate has been included in the list of Arts-related deletion discussions.
	No argument	Non-templated; meta-arguments	Officially withdrawing nomination. I didn't know if the above statement brought notice as far as withdrawing the nomination went, but I'm posting this just in case.	Comment. The same article appears to have been discussed a month and a half ago under a different title: Wikipedia:Articles for deletion/Transparent LED-embedded glass. That discussion resulted in 'keep'.

**Table 1. The 17 argumentation schemes annotated in round 4, with examples taken from our annotation manual. Argumentation schemes that can be problematic are marked; these are further discussed in RQ3 results.**

Cohen's Kappa	.48
Kappa Maximum	.75
Kappa Std Err	.01
weighted Kappa	.50
Percent Agreement	54%
Chance Agreement	12%

**Table 2. Inter-annotator agreement for Round 4 annotation**

with a standard parser, finally regular expressions were used to trim excess page content (including the consensus result and archiving notices) and to add linebreaks at the timestamps (providing a visual indication of message boundaries).

### Identifying Novices

After classifying arguments in the entire 72-debate corpus as described above<sup>15</sup>, we identified participants, establishing novices as those with either low overall edit count<sup>16</sup> (< 105) or no edits to deletion discussions prior to the observation period. We culled participants from user Talk page links in the HTML, then used the API to retrieve their contribution histories prior to January 14, 2011, the start of the first debate. Novices were confirmed manually. 22 double-novices<sup>17</sup> met both the low edit count and low discussion participation criteria. A further 6 participants met only one criterion (3 wiki novices had low edit count and 3 deletion novices had no previous participation in deletion discussions). The debates which included contributions from these 28 novices formed a subcorpus of 28 articles that was used to determine how an arguer's experience affects the arguments they give (RQ2).

### RESULTS

We present results for each of the research questions in turn.

#### Argumentation Schemes Used (RQ1)

Our core corpus of 72 debates yielded 1213 arguments in 741 messages. The most common argumentation schemes (RQ1) are shown in Table 3; each of these 14 schemes is used in at least 2% of arguments. Two non-argumentative categories, Note, and No Reason Given, also each comprise more than 2% of messages, as shown in Table 3.

Certain schemes are more prevalent compared to a general argument corpus: five of Wikipedia's top third most common deletion arguments—Arguments from Rules, Values, Bias, Precedent, and Waste—are not in the top two-thirds of the most commonly used arguments (pers. communication, Snaith) in the only widely-available informal argument corpus, the Araucaria corpus [21]. This shows Wikipedia's focus on precedent and rules, and the tendency to discuss articles in terms of both values or community norms and article contributors or supporters. Certain common schemes have no relevance to debates: for instance Fear Appeal and Distress were not observed, and Popularity was rare.

#### Differences in Argumentation: Novices vs. Experts (RQ2)

We next examined whether a participant's arguments depend on their experience with Wikipedia editing or the Wikipedia

<sup>15</sup>Novice classification took place after the second round of annotation; the third and fourth rounds of annotation was conducted by student annotators who were unaware of the classification.

<sup>16</sup>This was chosen based on a natural breakpoint in the contribution count of our user sample; previous research has indicated that about 100 edits is a moderately low edit count.

<sup>17</sup>Including 9 of 10 IP address 'users'; the tenth, though a wiki novice, contributed to 16 previous debates starting in October 2010.

deletion process. We focused on a subset of our core corpus, using only the 28 debates in which novices participated. We found 555 arguments, including 93 arguments in 46 comments from novices. This subset corpus showing novices' participation is further analyzed in Table 3, which shows how often novices, compared to everyone (novices+experts) used a particular argumentation scheme.

Argumentation schemes used by novices and experienced users are different (RQ2); statistically significant differences are given in bold in Table 3. Experience leads to more Arguments from Precedent and, unsurprisingly, more contributions of non-argumentative, administrative notes. Novices are more likely to give arguments which can have underlying flaws, using schemes such as Argumentation from Values, Argument from Cause to Effect, and Argument from Analogy, as will be detailed in RQ3 results.

The volume of participation is markedly different between novices and experts. In our corpus, only 6 novices made more than one comment, and only 3 novices participated in more than one debate (2,6,8 debates)(debates avg 2.2, stddev 1.9). By comparison, only 2 experts commented in just one of the 72 debates in our sample; overall, experts averaged 5.4 comments in 3.6 articles (stddev 10.1 for comments and 6.6 for articles). While the two largest number of comments (99 comments in 62 articles and 32 comments in 24 articles) were due to administrative notes and debate closing decisions, three experts participated in 11 to 20 debates and seven experts made 10 to 20 comments.

While regulars may read all debates, or all debates on a topic, novices are attracted to a debate by their knowledge of a topic, or by their desire to become more active in the community (for instance to support a nomination for adminship). Novices may be interested parties – creators or subjects – who may argue to keep an article without sufficiently understanding the criteria to be applied. Cases which attract many novices can become contentious, causing challenges unless there is strong policy to be applied [24]. In our sample, 'no consensus' discussions either lacked significant discussion (2) or involved discussions with a novice creator (2). Discussions with article creators showed strong emotion in five out of six cases, as indicated by the use of high sentiment words.

Novices arguing for deletion of an article are more likely to be participating in multiple arguments, and are more likely to cite policy correctly. Our sample showed a clear difference in the rhetoric of keep and delete comments, for example, in the ten comments left by IP users. Keep (6) comments exhibited emotional involvement and indicated confusion about policies, especially on notability, original research, and verifiability. Delete comments (4) used standard formatting and supporting resources. While further research is needed, this provides deeper insight on Lam et al.'s earlier findings that Keep outcomes involving novices are more likely to be overturned [24].

Experience with deletion has some effect on the skill and rhetoric with which people argue a case. In Wikipedia, policy knowledge is a particular stumbling block for novices. The difference between real-world importance and Wikipedia-

RQ2 More often given by	RQ3 Can be prob.	Argumentation Scheme	RQ1 Full corpus 72 debates (novices & experts)	RQ2 Subcorpus 28 debates (novices & experts)	RQ2 Subcorpus 28 debates (novices only)	RQ2 $\chi^2$	RQ2 Sig. level
		Argument from Evidence to Hypothesis	19.29%	20.65%	21.51%		
		Argument from Rules	16.90%	18.66%	12.90%		
<b>Experts</b>		<b>Note</b>	<b>13.69%</b>	<b>10.14%</b>	<b>0.00%</b>	$\chi^2=3.66$	p < .05
<b>Novices</b>	△	<b>Argumentation from Values</b>	<b>4.20%</b>	<b>4.17%</b>	<b>9.68%</b>	$\chi^2=5.13$	p < .05
		Argument from Need for Help	4.12%	3.62%	2.15%		
		Argument from Bias	3.87%	4.53%	6.45%		
		No reason given	3.22%	3.62%	5.38%		
		Argument from Position to Know	3.05%	2.36%	4.30%		
<b>Experts</b>		<b>Argument from Precedent</b>	<b>3.05%</b>	<b>3.80%</b>	<b>0.00%</b>	$\chi^2=3.66$	p < .1
	△	Argument from Ignorance	2.97%	3.08%	4.30%		
		Argument from Composition	2.56%	2.54%	2.15%		
<b>Novices</b>	△	<b>Argument from Cause to Effect</b>	<b>2.31%</b>	<b>2.72%</b>	<b>7.53%</b>	$\chi^2=5.59$	p < .02
<b>Novices</b>	△	<b>Argument from Analogy</b>	<b>2.23%</b>	<b>2.36%</b>	<b>8.60%</b>	$\chi^2=9.86$	p < .05
	△	Argument from Waste	2.23%	2.54%	1.08%		
	△	Practical Reasoning	2.23%	2.72%	3.23%		
		Arg. from Verbal Classification	2.06%	1.45%	2.15%		

**Table 3. The most common argumentation schemes, sorted by prevalence in our full corpus. Prevalence in the the subcorpus of 28 debates to which novices contributed is also shown, both for novices and everyone (novices & experts). Statistically significant differences between novices and experts' argumentation (based on Round 2 annotation) are noted in bold. For definitions and examples, see Table 2.**

importance (notability) is quite confusing to newcomers; this newcomer asks

*Why an article on Juvenile Justice System Rules prevalent (sic) in the largest province of Punjab having population of more than 90 million people including juveniles is not Notable?*

Among novices, confusion about reliable sources (RS) is particularly prominent—for instance claiming that a self-published biographical website counts as a source for a biography, or providing insufficient detail (e.g. *I see notability and RS*, without specifying the reliable sources). Verifiability, which is of particular importance for avoiding bias, can also be misunderstood; this newcomer wants to rely on plausibility instead:

*if you folks had been around actively working on the Web in 2000, you would know when the Dot Com Bubble burst, many, many companies went out of business. Servers with information about me... were taken off line.*

We further discuss problematic arguments next, looking at novices' and experts' arguments that are not compelling.

### Which Arguments are (not) Accepted (RQ3)

Newcomers are particularly likely to use certain common, but problematic, arguments, which are enshrined in policy and in essays such as Arguments to Avoid.<sup>18</sup> We describe some common problematic arguments, along with its Wikipedia nickname and its Walton scheme in Table 4.

#### Novices

Argumentation from Values is the second most common argument given by novices; this is unfortunate since it is problematic when it is used to argue from individual (rather than

accepted community values). Hence novice arguers sometimes provide unconvincing arguments, based on ignorance of community standards, e.g. to keep an article because it is *obviously of interest to the public in general*.

Similar problems with Argumentation from Values are found in the AfD for Emsworth Cricket Club. With 17 messages, Emsworth Cricket Club is one of the longest debates in our sample, yet could have been quickly decided. All four novices—two IP users and two with newly created accounts—advocate keeping, while experienced participants argue unanimously for deletion. While one novice argues well, pointing to the British Newspaper Library (though not specifically to any individual references), the other novices' comments are largely unconvincing, arguing from personal, but not shared, Wikipedia values (e.g. *Why just because it is a small team and not major does it not deserve it's own page on here?*). Delete comments, however, cite specific notability policies and the need for reliable secondary sources. The novices have not sufficiently understood the criteria to be applied, or the possible counterarguments. Such discussions can escalate into heated discussions which increase effort and may alienate participants, unless diffused by skillful arguing backed by strong policy.

In this case there is strong policy, but perhaps a lack of tact. The final word goes to an experienced participant who urges reading the policy, saying *This specialness (sic) argument is getting lame*. Socializing newcomers involves not only informing them about policy, but also taking an encouraging tone; in this case, warning away inexperienced (and hence presumably non-valuable) contributors takes priority over socializing them.

<sup>18</sup>[http://en.wikipedia.org/wiki/Wikipedia:Arguments\\_to\\_avoid](http://en.wikipedia.org/wiki/Wikipedia:Arguments_to_avoid)



Potentially problematic argument	Typical Argumentation Schemes	Wikipedia terminology
Personal preference	Argumentation from Values	ILIKEIT
Few search engine results	Argument from Ignorance	Google test
Large number of search engine hits	Argument from Cause to Effect	Google test
Requesting a favor	Argumentation from Values, Argument from Waste, or Practical Reasoning	PLEASEDONT
Analogy to other cases	Argument from Analogy	OTHER
No harm in keeping an article	Argumentation from Values or Argument from Waste	NOHARM
Topic will be notable in the future	Practical Reasoning	CRYSTALBALL

**Table 4. Some common problematic arguments, along the Walton scheme and the Wikipedia acronym.**

In good outcomes, novice creators learn policies and get mentoring on improving an article during the process: these fueled the No Consensus decisions in our sample<sup>19</sup>. Yet in other cases, novices who do not know the rules do not get mentored; they leave with the article deleted and no more insight into why the sources or notability were insufficient, with unanswered questions such as *If this page is deleted can it be added again once it has reliable sources?* In cases where experienced users are unanimous, there may be a particular incentive to ignore the voices of newcomers: this retains decision quality and minimizes effort at the expense of alienating participants.

In our corpus, the most severe conceptual errors were demonstrated by novices. Novices' comments may indicate lack of clarity on what a reliable source is (2); confusion about the relationship between reliable sources and notability (2); or failure to understand that reliable sources are needed for verification of a topic's existence (2). By comparison, experts tend to be familiar with policies and guidelines.

#### Experts

The problems demonstrated among experienced users, were significantly less severe. Sometimes such vagueness or rhetorical mistakes attract counter arguments. For instance *This article is inside WP:Bio<sup>20</sup> and should be kept* attracts a reply asking

*...which part of WP:Bio is met...actually providing some rationale would be more helpful than a WP:VAGUEWAVE<sup>21</sup>.*

In other cases, unchallenged arguments fail to be convincing. For example, a statement such as *It is an important historical part of Hamburg and I believe it deserves mention here.* would be decisive with sources. Such lack of justification or insufficient detail is common, yet detracts from arguments.

Even very frequent participants may make boilerplate arguments that lack enough detail to be convincing. For instance, out of 20 messages by one experienced user in our corpus, 15 say only *Delete – notability not demonstrated in a reliable secondary source.* This expert's 5 remaining messages argue for keeping an article (1); name what concept is not notable (1); and mention a second, additional concern (1); but only one provides an explanation of the problem (*I can't tell if any of this is notable stuff, because no sources tell us.*). The lack of justification is troubling because the argument that there are no reliable sources is an Argument from Ignorance: one

key critical question is whether a given search for sources was sufficiently thorough.

Among experienced users, there can be disagreements, for instance, about what "good sourcing" means or whether given sources are sufficient to establish notability. Counting sources is not sufficient: understanding what makes coverage appropriate is also needed. The fact that sources are used both to verify information and to establish a topic's importance complicates some arguments even amongst experienced commenters (e.g. *The reliability of the sources is not being questioned, nor is it an issue as far as this AfD is concerned. The issue is, and remains, notability.*).

Experts also make strategic errors in arguing: Four experts in our corpus present arguments based on an objection to policy (e.g. *Frankly, the basis of my disagreement with you here is that I don't agree with the guideline*), but these are unconvincing and do not carry the debate. One administrator's closing justification notes this pointedly, admonishing: "do not make up rationales contrary to policy and guidelines on individual AfDs"; rather, as the administrator notes, policies are to be developed on discussion pages. Essentially, the experts have made an error of jurisdiction.

## DISCUSSION & CONCLUSIONS

### Supporting Novices

Wikipedia policies are complex [14], and for novices, this poses problems of both learning and understanding the guidelines and policies. Newcomers would benefit from support in finding their way through the vast, dispersed information [4]. The deletion process can be confusing, and sometimes discouraging, especially to newcomers, who may become disillusioned or frustrated when content they contributed is deleted for reasons they don't always understand: Wikipedia's notability<sup>22</sup> policy, the main reason for up to 28% of deletions [25], is especially likely to be misunderstood by newcomers. Helping new arguers ground arguments in policy could be helpful, since rules and policy indicate the main consensus values; by contrast, statements grounded in personal values, such as *obviously of interest to the public in general* do not make a significant contribution to the outcome of a debate. Since users don't read instructions [6], to maintain quality in user-generated content without disillusioning users, summarizing content policies and presenting them in task-based contexts seems essential.

We also suggest increased sensitivity and attention to support involved parties—page creators and those who edited a page

<sup>19</sup>See [30] for detailed examples.

<sup>20</sup>[http://en.wikipedia.org/wiki/Wikipedia:Notability\\_\(people\)](http://en.wikipedia.org/wiki/Wikipedia:Notability_(people))

<sup>21</sup>[http://en.wikipedia.org/wiki/Wikipedia:Arguments\\_to\\_avoid\\_in\\_deletion\\_discussions#Just\\_pointing\\_at\\_a\\_policy\\_or\\_guideline](http://en.wikipedia.org/wiki/Wikipedia:Arguments_to_avoid_in_deletion_discussions#Just_pointing_at_a_policy_or_guideline)

<sup>22</sup><http://en.wikipedia.org/wiki/WP:Notability>

before it was nominated for deletion. The emotional attachment of this group adds complication, especially when they are new to the deletion discussion process. Based on our corpus, newcomers are more likely to take the process personally, making statements such as *I know wikipedia (sic) has a dislike for all things [article topic]*. In contrast, experienced Wikipedians' contributions do not typically show high levels of emotional involvement. Rather, experienced Wikipedians appear to use personal pronouns to report on actions they have taken (e.g. changes made to the article), to suggest new policies, to mark uncertainty (e.g. *I would say*), or to show a personal opinion that may differ from the consensus view (e.g. *I don't expect my saying that to alter the outcome of this AfD*). Statistical analysis of the use of sentiment and personal pronouns could provide further evidence for the differences in personal engagement we see. Yet we think that these differences are sufficiently clear to warrant immediate community response in attending skillfully to newcomers.

The emotional tenor of discussions can change dramatically based on the skillfulness and sensitivity of the responses newcomers receive. For example, one newcomer from our corpus, now a prolific contributor, was initially frustrated when their first article was nominated for deletion, commenting: *To be honest it's been a real turn off adding articles to WP and I don't think I will add articles again. So smile and enjoy*. It appears that the support and co-editing from other editors during the deletion process tempered this frustration. Balancing emotion and detachment seems helpful, as exemplified by this quote, from another novice editor finally accepting that deletion is the right outcome: *I believe that (much as it would break my heart based on the no of hours I have put in over the years working on the article) it is perhaps sensible that the piece is deleted*. Understanding the process and what made it *sensible*, following extensive iterative discussion with more experienced editors, appears to have tempered the emotion for this editor. Feedback and mentoring during the AfD can impact the outcome for editors, potentially mediating the negative emotional impact. Existing work on values expressions in Wikipedia discussions [26] and on language use for mediation [3] may provide some guidance; further work, more specific to deletion, is needed. For instance, a small fraction of new editors do continue editing after an article they created was deleted; further investigation of their motivations and experience could be helpful.

### Supporting All Participants

In future work we are developing experimental interventions for deletion discussions in hopes of defusing contentious conversations while also supporting newcomers in learning the relevant policies and rhetoric. For instance, extracting discussion topics could help focus the discussion on the disputed points. People often explicitly identify what they believe the issues to be, in attempts to focus the discussion. Standardized, easily detected phrases may be used, such as, 'the real issue is:' or 'I haven't heard any arguments yet which refute that point. In addition, coordination, social, and instructional messages can also appear in discussions, for instance to thank someone else; to point out a change a user just made; or to provide instructional content. We would like to test whether

these interventions matter to newcomers' ongoing community participation as well as to the community satisfaction with the decision.

To develop future support mechanisms, focusing on argument structure could also be helpful. Walton's schemes can be construed as one approach to quality measurement, allowing arguments to be weighed against one another, and their strengths compared. In particular, each argumentation scheme indicates the critical questions that attack the argument; this can help to decide which argument prevails. To structure argumentation based on community standards, community-specific templates based partially instantiating the most common argumentation schemes could be used. Use of argument templates might simplify the structure of typical debates while aiding participants in making appropriate arguments, potentially bring both rigor and ease.

### Studying Social Sensitivity in Online Ad-Hoc Task Groups

While there is existing research on dedicated arbitration in text-based arguments [3], further research is needed on the collective intelligence of online *ad-hoc* task groups. For in-person task groups, collective intelligence correlates with the average social sensitivity of a group [35]; if similar phenomena operate online, techniques for increasing the social sensitivity of ad-hoc online task groups would be helpful.

Social sensitivity and neutrality may also help counter the fatigue of ongoing participation in contentious activities such as deletion. A taxonomy of the emotional triggers and the associated needs discussants are trying to address (e.g. 'understand why this article was deleted', 'provide further information about a point that was not taken into consideration', 'vent about policy and bureaucratic challenges') could help suggest ways to meet participants' needs while defusing emotional debates, and might suggest likely subproblems that could be fruitfully addressed.

### Reusing Argumentations Analysis Methods

The analysis techniques we describe can also be reused. This process is intensive and difficult to apply, yet it is appropriate to and transferrable to any textual discussion or online system where it would be valuable to understand the argumentation at a high level of detail. This would allow a comparison of the reasoning patterns and argumentation used.

Understanding the argumentation used in online discussions throughout the Web is also important. Towards a deeper understanding of persuasive and acceptable argumentation in online fora, Walton's argumentation schemes—which he acknowledges are not exhaustive—should be customized to particular domains (e.g. online reviews, open source bug reports) and fora (e.g. Wikipedia deletion discussions, W3C standardization discussions). This would enable support for newcomers, an understanding of common fallacies in different genres of Web conversations, and a comparison of the rhetoric. For instance, online fora make different uses of policy: In the context of Wikipedia, appropriate rhetoric and good use of policy are closely connected to successful arguments. We expect different results in communities with a less well-developed policy sphere. Application of the analysis techniques used in this paper is a first step towards that deeper understanding.

## Acknowledgements

This work was supported by Science Foundation Ireland under Grant No. SFI/09/CE/I1380 (Líon2). Thanks to the Wikipedia community, especially those we informally interviewed about deletion, and its researcher community. We are particularly grateful to Bernie Hogan, for pointing out that deletion is of wider interest than article discussions. For annotation work, we thank Laura O'Connor and Lyndia Peters. We thank Mark Snaith for data on argument prevalence and Adam Wyner for conversations on argumentation. We also thankfully acknowledge the anonymous reviewers for several rounds of valuable feedback starting with the 2012 revise and resubmit process, and comments from Luigina Ciolfi, David Randall, and Trevor Bench-Capon on an early draft.

## REFERENCES

1. J. Antin and C. Cheshire. Readers are not free-riders. In *CSCW '10*.
2. F. Barcellini, F. Détienne, and J.-M. Burkhardt. Participation in online interaction spaces. *Int'l Journal of Industrial Ergonomics*, 39(3):533–540, 2009.
3. M. Billings and L. A. Watts. Understanding dispute resolution online. In *CHI '10*.
4. S. L. Bryant, A. Forte, and A. Bruckman. Becoming Wikipedian. In *GROUP '05*.
5. B. Butler, E. Joyce, and J. Pike. Don't look now, but we've created a bureaucracy. In *CHI '08*.
6. J. M. Carroll and R. L. Mack. Human-computer interaction. In *Learning to use a word processor: by doing, by thinking, and by knowing*, pages 698–717. Morgan Kaufmann, 1995.
7. P. K. Chilana, A. J. Ko, and J. O. Wobbrock. Understanding expressions of unwanted behaviors in open bug reporting. In *IEEE VL/HCC '10*.
8. B. Choi, K. Alexander, R. E. Kraut, and J. M. Levine. Socialization tactics in Wikipedia and their effects. In *CSCW '10*.
9. K. Crowston, Q. Li, K. Wei, U. Y. Eseryel, and J. Howison. Self-organization of teams for free/libre open source software development. *Information and Software Technology*, 49(6):564–575, June 2007.
10. W. Emigh and S. C. Herring. Collaborative authoring on the web. In *HICSS'05*.
11. S. Farnham, H. R. Chesley, D. E. McGhee, R. Kawal, and J. Landau. Structured online interactions. In *CSCW '00*.
12. R. Farzan, R. Kraut, A. Pal, and J. Konstan. Socializing volunteers in an online community. In *CSCW '12*.
13. A. Forte, N. Kittur, V. Larco, H. Zhu, A. Bruckman, and R. E. Kraut. Coordination and beyond. In *CSCW '12*.
14. A. Forte, V. Larco, and A. Bruckman. Decentralization in Wikipedia governance. *Journal of Management Information Systems*, 26(1):49–72, 2009.
15. R. S. Geiger and H. Ford. Participation in Wikipedia's article deletion processes. In *WikiSym '11*.
16. R. S. Geiger, A. Halfaker, M. Pinchuk, and S. Walling. Defense mechanism or socialization tactic? *ICWSM '12*.
17. A. Halfaker, R. S. Geiger, J. Morgan, and J. Riedl. The rise and decline of an open collaboration system. *American Behavioral Scientist*, In Press.
18. E. Hovy. Annotation tutorial. In *ACL '10*.
19. M. Huysman, C. Steinfield, C.-Y. Y. Jang, K. David, M. H. I. N. T. Veld, J. Poot, and I. Mulder. Virtual teams and the appropriation of communication technology. *CSCW*, 12(4):411–436, 2003.
20. E. Joyce, B. Butler, and J. Pike. Handling flammable materials. *iConference '11*.
21. J. Katzav, C. A. Reed, and G. W. A. Rowe. Argument research corpus. *Practical Applications in Language and Computers '03*, pages 229–239.
22. S. W. Kozlowski and D. R. Ilgen. Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest*, 7(3):77–124, 2006.
23. R. Kraut, M. Burke, J. Riedl, and S. Drenner. “Dealing with newcomers”. In *Evidence-based social design*. MIT Press, 2010.
24. S. K. Lam, J. Karim, and J. Riedl. The effects of group composition on decision quality in a social production community. In *GROUP '10*.
25. S. T. K. Lam and J. Riedl. Is Wikipedia growing a longer tail? In *GROUP '09*.
26. J. T. Morgan, R. M. Mason, and K. Nahon. Lifting the veil: the expression of values in online communities. In *iConference '11*.
27. K. Nemoto, P. Gloor, and R. Laubacher. Social capital increases efficiency of collaboration among Wikipedia editors. In *HyperText '11*.
28. I. Rahwan. Mass argumentation and the semantic web. *Web Semantics*, 6(1):29–37, 2008.
29. C. Reed. Walton's theory of argument and its impact on computational models. In C. Reed and C. W. Tindale, editors, *Dialectics, Dialogue and Argumentation*. College Publications, London, 2010.
30. J. Schneider, A. Passant, and S. Decker. Deletion discussions in Wikipedia: Decision factors and outcomes. In *WikiSym '12*.
31. B. Suh, G. Convertino, E. H. Chi, and P. Pirolli. The singularity is not near. In *WikiSym '09*.
32. D. Taraborelli and G. L. Ciampaglia. Beyond notability. In *IEEE SASOW '10*.
33. D. Walton. *Argumentation Schemes for Presumptive Reasoning*. Routledge, 1995.
34. D. Walton, C. Reed, and F. Macagno. *Argumentation Schemes*. Cambridge, 2008.
35. A. W. Woolley, C. F. Chabris, A. Pentland, N. Hashmi, and T. W. Malone. Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004):686, 2010.