



Add image:
User testing findings
(Round 1 - Dec 2020 EN)

Rita Ho & Marshall Miller | Jan 2021

Background

Multiple product teams—Android, Growth, and Structured Data—are working on a new [structured task](#) (f.k.a. microcontributions) enabling users to evaluate images that have been suggested by an algorithm to be added to unillustrated Wikipedia articles.

This research sought initial feedback on the quality of the image suggestions algorithm in its current state (as of Dec 2020); as well as to better understand to what extent audiences unfamiliar with editing on Wikipedia can complete tasks.

Additional product context

The image suggestion algorithm tested has been created by Miriam Redi on the Research Team (see [T266271](#)). Information from this study may be useful in validating and suggesting improvements to the algorithm; as well as provide insight into which metadata users need to complete tasks.

The test prototype's design follows the shared “task feed” format already in use in the respective “Suggested edits” feature in the Android app and the Growth team's newcomer homepage.

Target audience

The focus for this test were people with limited experience editing on Wikipedia or Commons. Additionally, as Android will likely be the first product team to utilise this tool, tests were conducted on mobile.

In summary, participants characteristics were the following:

- **Experience contributing:** Few or no edits to Wikipedia or Commons.
- **Platform:** Mobile.
- **Language:** English for the first round of testing, but we intend to replicate the study with participants in another language who are not bilingual in English to see how the limited understanding of the large proportion of English image metadata may affect results.

Research goals | Questions

Primary objective

Evaluate task feasibility and attractiveness for newer contributors.

1. Are participants able to confidently confirm matches based on the suggestions and data provided?
2. How accurate are participants at evaluating suggestions? And how does the actual aptitude compare to their perceived ability in evaluating suggestions?
3. How do participants feel about the task of adding images to articles this way? Do they find it easy/hard, interesting/boring, rewarding/irrelevant?

Secondary objectives

Evaluate information quality.

4. What information do participants find most valuable in helping them evaluate image and article matches?
5. Are participants able to write good captions for images they deem a match using the data provided?

Identify opportunities for usability improvements.*

6. Were there any pain points in the display of and navigation to information that hindered participants' ability to make matches?

*Note: The prototype design is an early, barebones execution as a fluent user experience is not the focus of this study.

Participants & Test format

An unmoderated remote task-based test was conducted on [UserTesting.com](https://www.usertesting.com).

Fifteen^[1] people participated in total, tested in three rounds of 5 people (denoted Group A, B, and C)^[2].

Respondents were screened for those less familiar with editing on Wikipedia.

[1] See NNGroup article on user testing sample sizes:
<https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>

[2] For Group B, minor changes were made to the UI and task instructions to highlight the article card (as some Group A testers did not immediately make the connection). For Group C, the same 60 images tested in Groups A and B were shown in reverse order. Minor copy changes to the article card label. See the appendix slides for more details on differences across groups.

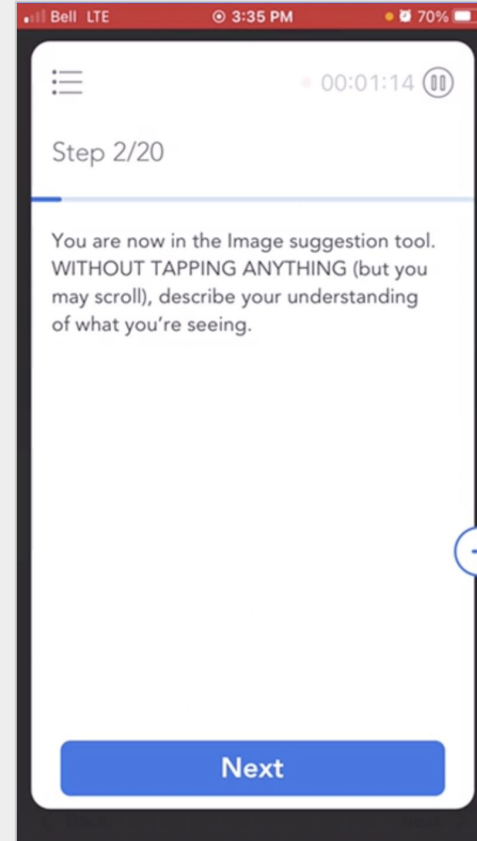
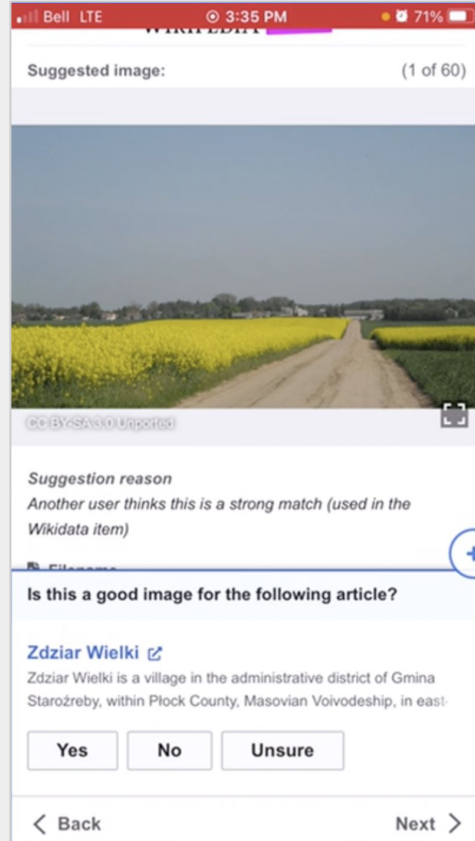
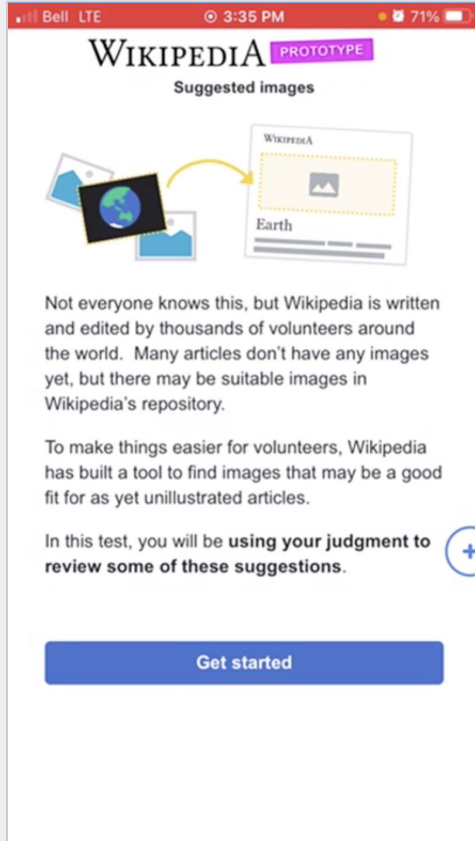
Participant demographics

- **Gender:** 9 Female, and 6 Male.
- **Age:** From 20-60 years old.
(6x 18-35yo, 6x 35-50yo, 3x 50+)
- **Countries:** 8 USA, 3 Canada, 2 UK, and 2 Italy.
- **Platform:** 11 iOS, and 4 Android
- **Wikipedia usage:** Daily to Monthly usage
1x “Multiple times a day”, 7x “Daily”, 4x “2-3 times a week”, 1x “Weekly”, 2x “2-3 times a month”.
- **Have edited Wikipedia:** All had little or no edits.
 - 5x “No, because I didn't know it could be edited”
 - 4x “No, for other reasons”
 - 6x “Yes, once or twice”
- **Have used Wikimedia Commons before:** 3 people.
3 selected this in the screener “Which of the following sites have you used before?”, (vs 5 who selected Flickr as well.)

Main tasks in the study

- Review at least 20 of the 60 image suggestions presented to them. Participants were asked to explain their decision-making out loud.
- Explain what they think the caption dialog is about, and then write a caption for at least one image suggestion.
- Rate the usefulness of different information about the image and article, in helping them to complete the image matching task.

Tasks overview - Part 1: Intro/Orientation



Initial tasks giving context that this tool is to help add images to unillustrated Wikipedia articles.

Users are asked to give their first impressions after the introduction, to help gauge their broad understanding.

Tasks overview - Part 2: Review images

Bell LTE 3:37 PM 70%

☰ 00:02:51

Step 4/20


Please go through these suggestions and review them as best as you can. **TAP AND INTERACT** with them as you would in real life, while also **TALKING OUT LOUD** your thought process. Specifically, **how are you deciding** on whether to choose Yes, No, or Not sure?

Note: You don't have to do all 60, but aim to answer at least 20 before moving to the next task. You can tap on the article title to open the full article in a new window if this helps. Tap "Next" if you want to skip a suggestion entirely.

Next

Bell LTE 3:43 PM 67%

Suggested image: (6 of 60)



Suggestion reason
Used in the same article on another language Wikipedia:
Hebrew

Filename
Kollelbirkatizhak.jpg

2496 x 1664 pixels
22 Aug 2010
Birkatizhak
[More info \(Commons file page\)](#)

Image description
English: kollel "Birkat Itzhak" (Moscow) - study

Image caption
-

Categories
Jewish people of Russia, People learning Torah, Judaism in Moscow

Tags

Is this a good image for the following article?

[Kollel](#)

A kollel or colel is an institute for full-time, advanced study of the Talmud and rabbinic literature. Like a yeshiva, a kollel features

Yes No Unsure

< Back Next >

Bell LTE 3:43 PM 67%

Suggestion reason
Used in the same article on another language Wikipedia:
Hebrew

Filename
Kollelbirkatizhak.jpg

2496 x 1664 pixels
22 Aug 2010
Birkatizhak
[More info \(Commons file page\)](#)

Image description
English: kollel "Birkat Itzhak" (Moscow) - study

Image caption
-

Categories
Jewish people of Russia, People learning Torah, Judaism in Moscow

Tags

Is this a good image for the following article?

[Kollel](#)

A kollel or colel is an institute for full-time, advanced study of the Talmud and rabbinic literature. Like a yeshiva, a kollel features

Yes No Unsure

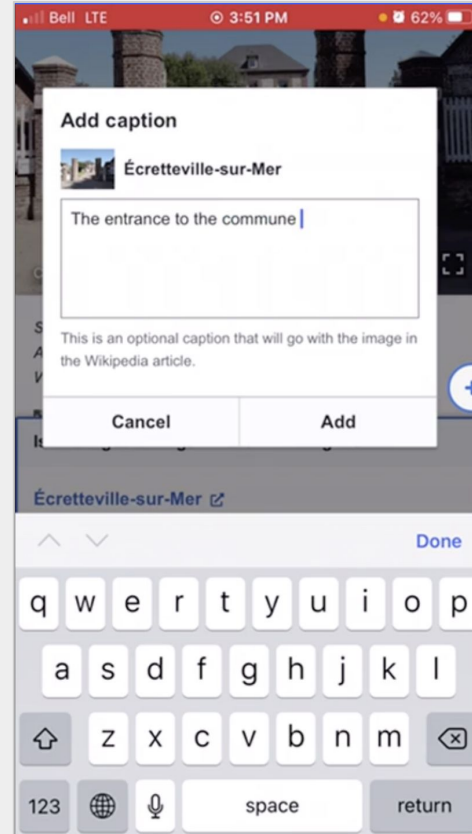
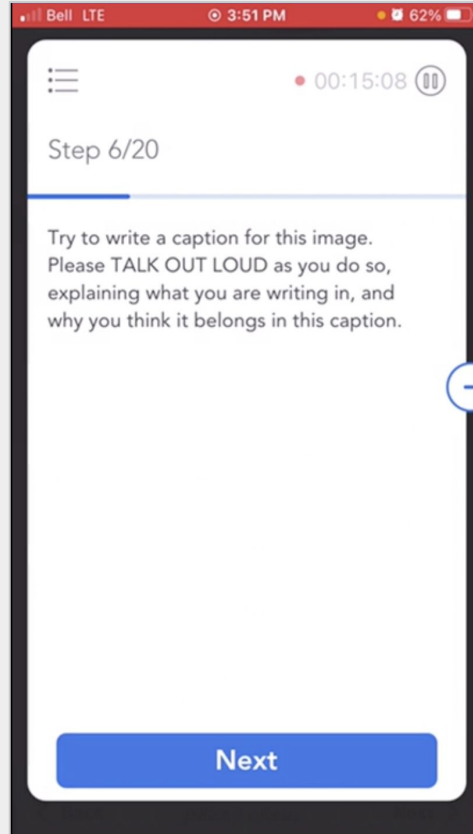
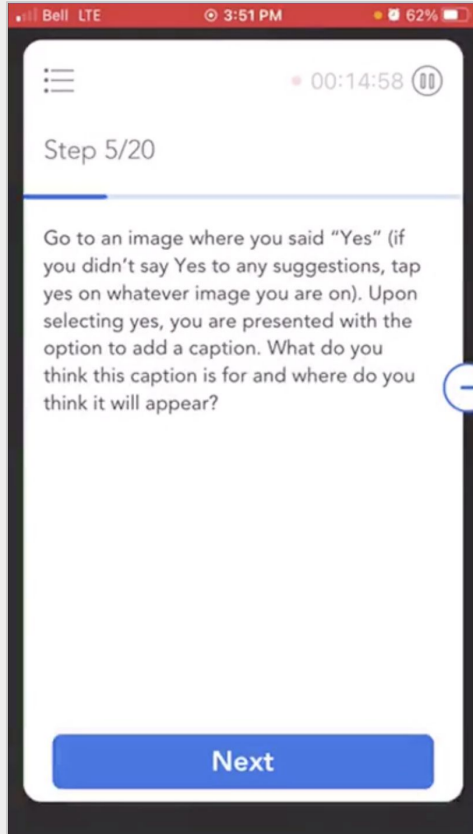
< Back Next >

Each image is presented with the same data (Image filename, image size, upload date, image description, etc.).

Some notes on the info:
Suggestion reason – the source of the suggestion from the image recommendation algorithm – it can be one of 3 reasons (1. image was in the Wikidata item, 2. image was on a category of the Wikidata item, or 3. image was used in the same article on another language or project).

Tags – the structured data “depicts tags” for the image.

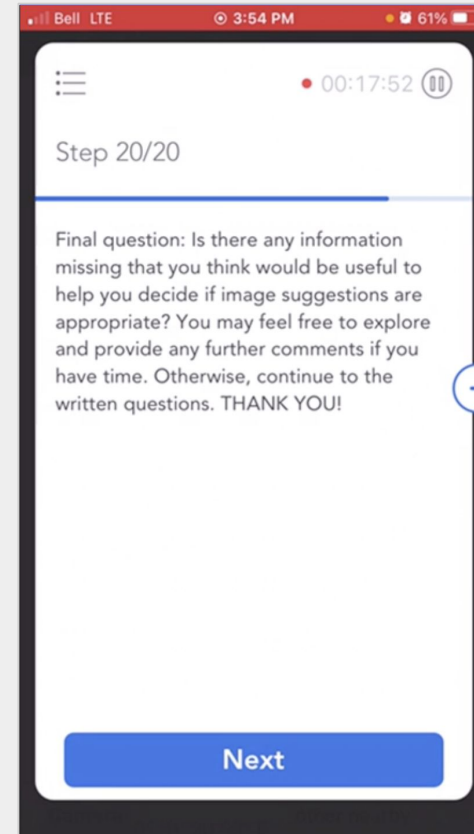
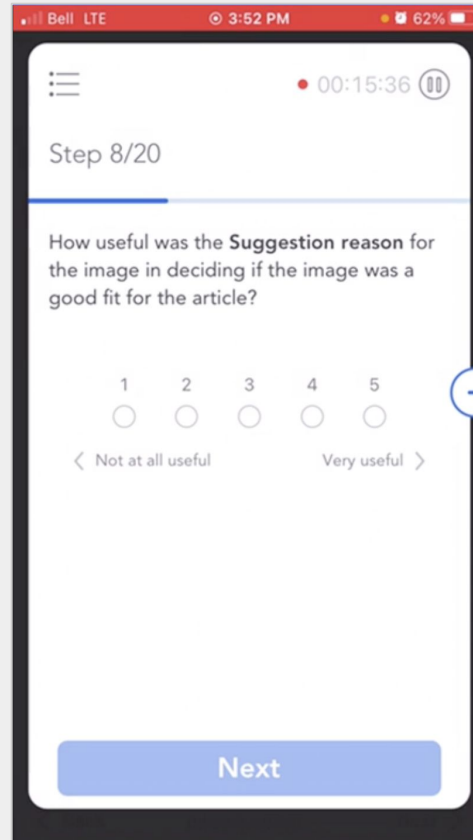
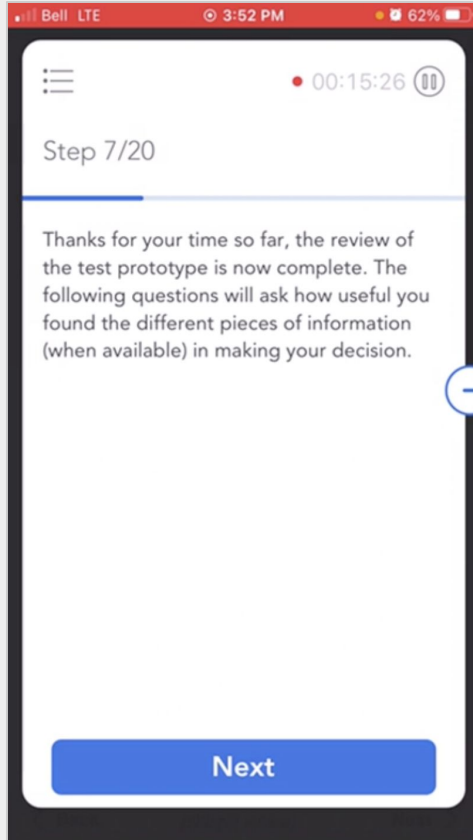
Tasks overview - Part 3: Caption



Participants are asked about their understanding of the caption they're prompted to add after accepting an image suggestion.

After answering where they think the caption appears and what they think it's for, they are asked to demonstrate writing a caption.

Tasks overview - Part 4: Rate information



Participants are asked to rate each piece of information from 1 to 5, before being asked if they felt anything was missing that may have helped in their decision-making.

Written responses after the study

1. How easy or hard did you find this task of reviewing whether images suggested were a good match for articles? (*Gauge perceived difficulty*)
2. Would you be interested in adding images to Wikipedia articles this way? Please explain why or why not. (*Gauge interest/engagement*)
3. Was there anything that you found frustrating or confusing, that you would like to change about the way this tool works? (*Usability improvements*)
4. How do you think the suggested images for articles are being found? And how would you rate the overall quality of the suggestions? (*User's quality perceptions based on trust/understanding of suggestions*)

Findings & Recommendations



Key takeaways

- **General understanding of the task matching images to Wikipedia articles was reasonably good**, given the minimal context provided for the tool and limited knowledge of Commons and Wikipedia editing. There are opportunities to boost understanding once the tool is redesigned in a Wikipedia UX.
- **Each image matching task could be done quickly by someone unfamiliar with editing**. On average, it took 34 seconds to review an image.
- **All said they would be interested in doing such a task**, with a majority rating it as easy or very easy.
- **Perceived quality of the images and suggestions was mixed**. Many participants focused on the image composition and other aesthetic factors, which affected their perception of the suggestion accuracy.
- **Most of the image metadata fields presented were self-rated as very useful for decision-making**. But in practice, it was observed that only a few pieces of data were critical for image matching.
- **Being able to view the whole article was highly rated**, despite few actually doing so in the test.
- **Many participants would, at times, incorrectly try to match images to the its own data, rather than to the article**. Layout and visual hierarchy changes to better focus on the article context for the image suggested should be explored.
- **Adding captions was poorly done**. It may be partly due to some initially misunderstanding the caption as being part of justifying their image match. However, even when a tester understood its purpose, caption content and writing style was generally mediocre.

Conceptual understanding

Do people know what it means to add images?

- **High understanding of the main task (evaluating whether a suggested image matched to an article).**
 - 9 testers knew from the start to review by comparing images and metadata to the article.
 - 4 others took a little time to make the right connection (sometimes as they initially missed seeing the article card, or misread the filename or other image info as the article).
 - 2 others were consistently ignoring the article, with one particular poor performer reviewing all images against its own metadata.
- **Poor understanding of where the images are from.**

Even though test instructions explicitly explains this, and there are links out to the “Commons file page”, no one mentions Commons. Instead, Google, Google images, or a “search engine” were assumed to be the source of images and/or suggestions.
- **The placement of the image and its metadata above an article preview seemed to confuse some testers.**

It took a while for some testers to notice and connect the article card. Others lost focus during the test and would sometimes match an image with its own metadata in error instead of the article.

 - **Action:** Consider revising the UI with the image suggestion more “in context” of the full article
 - **Action:** Incorporate a more ‘in-context’ article preview when participants select Yes.
 - **Action:** Provide onboarding screens with clear task guidance and examples.
- **Two participants assumed it was a tool for people during article creation.** Their assumption hits upon a future opportunity to incorporate it during editing.

Image evaluation

Task summary

Participant responses were compared to the actual correct answers* to measure the quality of their ratings.

There were three categories of actual correct answers:

- **Yes**
The image is a good match to the article
- **Yes (conditional)**
The image is a good match with conditions, requiring (a) research beyond the information shown, and/or (b) a caption for context.
- **No**
The image is not a good match

*Actual correct answers as evaluated by MMiller and RHo, the test authors.



Suggestion reason

Another user thinks it's related (image is in a Commons category linked to the article's Wikidata item)

Filename

Opatalj – Crkva Sveti Roka - 01.jpg

3456 x 2304 pixels

03 Sep 2013

Maesi64

[More info \(Commons file page\)](#)

Image description

English: Church of Saint Roch (Crkva Sv. Roka/ Chiesa di San Rocco), in Opatalj

Image caption

-

Categories

Saint Roch church (Opatalj), Saint Roch churches in Croatia, Built in 1470s churches in Croatia, Built in 15th-century churches in Croatia, Churches built in 1581

Tags

-

Is this a good image for the following article?

[Church of Saint Roch](#)

Church of Saint Roch or Saint Roch's Church may refer to:

Yes

No

Unsure

< Back

Next >

Image evaluation

Examples of different “actual” correct answers



GNU Free Documentation License, v1.2 or later

Suggestion reason

Another user thinks this is a strong match (used in the Wikidata item)

Filename
Guarda Veneta Town Hall.jpg

Size 3968 x 2976 pixels
Date 11 Dec 2010
Category Threcharlie

Is this a good image for the following article?

[Guarda Veneta](#)

Guarda Veneta is a comune (municipality) in the Province of Rovigo in the Italian region of Veneto, located about 60 kilometr

Yes No Unsure

< Back

Next >

Yes

Clear match of metadata (filename, description, category) and reason (Wikidata item)



CC BY-SA 3.0 Germany

Suggestion reason

Another user thinks this is a strong match (used in the Wikidata item)

Filename
Strassenfest in der Dreimuehlenstrasse in Muenchen 03.jpg

Size 512 x 335 pixels
Date 19 Apr 2017

Is this a good image for the following article?

[Diversity \(politics\)](#)

Diversity as seen in sociology and political studies is the degree of differences in identifying features among the members of a

Yes No Unsure

< Back

Next >

No

Generic image not representative of the subject.



GNU Free Documentation License, v1.2 or later

Suggestion reason

Used in the same article on another language Wikipedia: German

Filename
Triumph-tr2.jpg

Size 1000 x 586 pixels
Date 28 Feb 2006
Category Liftarn

Is this a good image for the following article?

[Ken Richardson \(racing driver\)](#)

William Kenneth "Ken" Richardson was a British racing and test / development driver from England who competed in one Formul

Yes No Unsure

< Back

Next >

Yes (conditional: research)

Need to read the dewiki article to know the driver invented the car, and then include this in the caption.



CC0

Suggestion reason

Used in the same article on another language Wikipedia: Spanish

Filename
Poliedro de Caracas Venezuela2.JPG

Size 921 x 684 pixels
Date 07 Oct 2013
Category Ángel Guerra Shifrin

Is this a good image for the following article?

[Miss Venezuela 2013](#)

Miss Venezuela 2013 was the 60th edition of the Miss Venezuela pageant held on October 10, 2013 at the Poliedro de Caracas ir

Yes No Unsure

< Back

Next >

Yes (conditional: caption)

Caption would need to note this stadium is where “Miss Venezuela 2013” was held.

Image evaluation

Algorithm accuracy

The 60 images used come from a [data set of ~14K](#) pulled from the Dec 2020 iteration of the algorithm.

In assigning the “correct answer” by which to compare participant responses, the algorithm accuracy for the sample data set was determined* as conservatively 53%, or 70% if including “Yes (conditional)” responses.

*NOTE: The limited sample size means this information is indicative and not statistically significant.

Correct answer	Count
No	18
Yes	32
Yes (conditional)	10
Grand Total	60

Algorithm accuracy	53%
Algorithm accuracy w/ "Yes (conditional)"	70%

Image evaluation

Algorithm accuracy by source/suggested reason

The images were selected at random, but were stratified in a similar ratio of different “Suggested reasons” as the full 14K dataset.

- **Interlink** (image was used in another language/project): 25
- **Wikidata** item (contained the image): 20
- **A Commons category** in the Wikidata item contained the image: 15

In reviewing this data, the algorithm’s author believes there is an order of algorithm accuracy by suggested reason to be as seen in the table to the right:

The figure displays three examples of image evaluation results, each with a numbered orange box (1, 2, or 3) indicating the suggested reason. Example 1 shows a dirt road through a field of yellow flowers with a license of CC BY-SA 3.0 Unported. Example 2 shows a page of handwritten text with a license of Public domain (United States). Example 3 shows a map of the Philippines with a license of Generic. Each example includes a 'Suggestion reason' text box.

1
CC BY-SA 3.0 Unported
Suggestion reason
Another user thinks this is a strong match (used in the Wikidata item)

2
Public domain (United States)
Suggestion reason
Used in the same article on another language Wikipedia:
French, Catalan, Scottish Gaelic, Danish

3
Generic
Suggestion reason
Another user thinks it's related (image is in a Commons category linked to the article's Wikidata item)

Accuracy by Suggested reason:

1. Wikidata item
2. Interlink
3. Commons category

Image evaluation

User accuracy

Overall scores indicate* newcomers would be able to use such a tool to successfully add a fair amount of images to articles (63% true positive).

The expected revert rate (17%) is within the normal range of new editor reverts, and potentially lower for image edits.

That said, this revert rate may be higher than communities will be comfortable with given the high visibility of images in an article.

Captions will likely affect the revert rate too. If the users are good at writing captions, more “Yes (conditional)” will survive. Bad captions may lead to fewer “Yes” surviving.

For this limited data set, removing an outlier poor performer (Brianna19) decreases the false positive and revert rates by 4pp and 2pp respectively. **This highlights the potential damage that a single contributor can make.**

*The limited sample size means this data is indicative and not statistically significant.

Correct answer	No	Skipped	Unsure	Yes	Grand Total
No	52	5	22	33	112
Yes	31	5	48	141	225
Yes (conditional)	30		14	18	62
Grand Total	113	10	84	192	399

Images that would be added to articles: 192

Number that might be reverted: 33

True positive rate 63%

False positive rate 29%

Revert rate 17%

Strict revert rate ^ 27%

^ Strict revert includes "Yes (conditional)" items

When the outlier participant is excluded:

False positive rate 25%

Revert rate 15%

Strict revert rate ^ 23%

Image evaluation

Interaction times and patterns

The time spent evaluating each image being on average 34s* provides a good indication that the task is within the expectations of an **easy** and short task, even for people unfamiliar with editing on Wikipedia.

The written responses reflect this, with 12 of 15 testers saying they found the task easy or very easy.

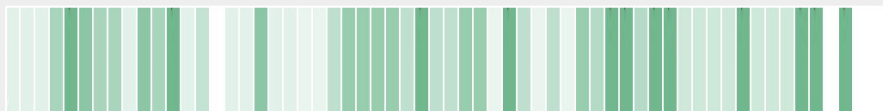
Importantly, as well as the majority self-reviewing the task as easy, the average agreement across participants is ~76%. This means that, on average, 3 out of 4 users agree that an image should either be a “Yes” or a “No. **This is considered a healthy magnitude for this number.**

*NOTE: The limited sample size means this information is indicative and not statistically significant.

Metrics: Review image task	Mean	Median
Number of images reviewed	26.1	20
Time evaluating each image	36s	34s

How much did participants agree in their decisions?

Each bar denotes an article reviewed (L→R most to least participants)



Average 76% agreement

Legend:

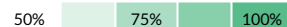


Image evaluation

1 of 3 – Detailed findings and observations on participants' aptitude & interest

- Many participants would match images to image data in at least part of the test. Sometimes people started the test not noticing the article card; whilst others would forget and myopically focus on the data closer to the image mid-test.
 - **Action:** Explore changes to layout and visual hierarchy so that the article context is more prominently focused.
- “Streaks” of good matches made some participants more complacent with accepting more images. This behaviour was anecdotally seen more with "Yes" than "No" matches.
 - **Action:** Explore introducing a range of difference confidence matches into a queue.
 - **Action:** Explore how to show image tasks in “sets” to break up repetitive interactions.
- Participants were often *overly* unequivocal in their decisions, under-using “Unsure” and skip. In some cases where the person had said out loud they are not sure, they nevertheless will still click Yes/No.
 - **Action:** Potentially provide more consequence to “No” such as asking for the rejection reason.
 - **Action:** Similarly provide review or confirm step when people select yes.
 - **Action:** Include quality metrics/gates to encourage more considered decision-making.
- Images in other languages (either on the image or its metadata) were prone to more rejection. People are hesitant about adding unfamiliar language content. Highly relevant in considering how well this tool may be used on its own by **non-english users**, since the majority of metadata on Commons is English only.

Image evaluation

2 of 3 – Detailed findings and observations on participants' aptitude & interest

- **Aesthetics/ Image quality judgement** play a part in decision-making. For example, an image of a specific moth species from the Wikidata item was rejected for being a “terrible picture”, whilst another correct image was rejected for being blurry (“I don't like that photo, it's fuzzy”).
 - **Action:** Provide accessible guidance and/or reminders about evaluating from the lens of suitability to the article.
- **Format bias: Photographs tended to be favoured more than illustrations/diagrammatic files** – some hesitated as to whether drawings (of animals) or diagrams (maps) were suitable.
 - **Action:** Provide examples and guidelines of acceptable media, compositions, etc.
- **Composition specificity was another ambiguating factor.** Particularly with biography or animal articles, testers were often unsure how “close-up” the subject needed to be, or it's necessary to crop or otherwise draw focus to the subject. E.g., For the article about a skier, the image shows the skier with her sister.
 - **Action:** Guidance for adding captions that specify the subject depicted
 - **Action:** Explore how basic image editing (e.g., [CropTool](#)) may be added as options in future

Image evaluation

3 of 3 – Detailed findings and observations on participants' aptitude & interest

- “Main” image as a factor in the final decision. Some users who knew that an image was related still said no based on their judgement that it needed to be more overtly connected or specific.

E.g. 1, for the article “[Miss Venezuela 2013](#)” the image of a basketball game in a stadium where the event was held was rejected by a couple of people who deemed it not related enough.

E.g. 2, the Scottish cup image was rejected by a tester for being “too vague” for someone new looking for (search) results

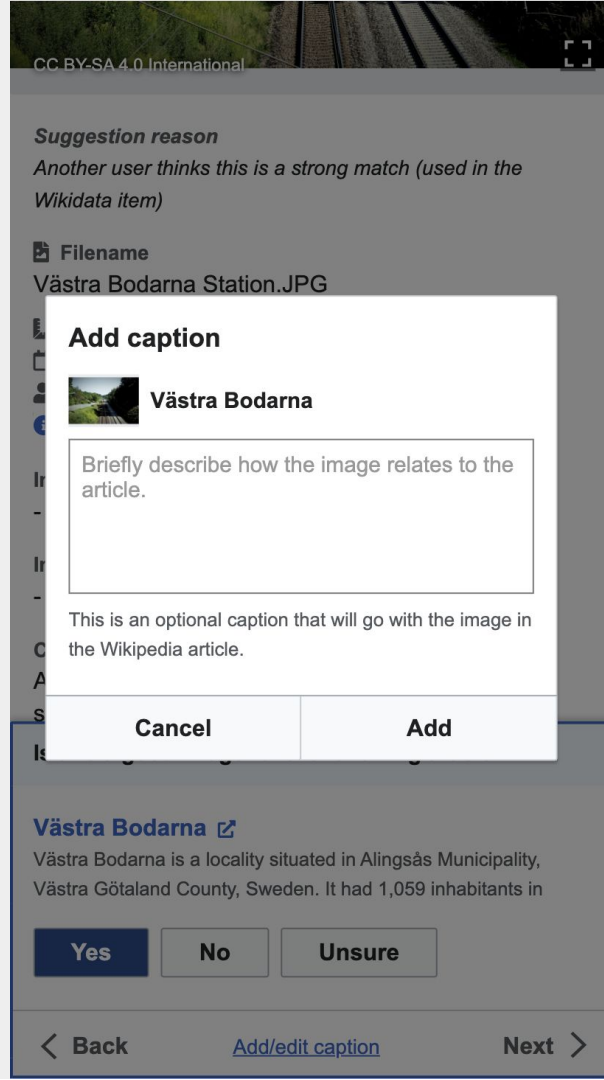
- Assumptions that other image choices are available may have played a factor for a couple of participants who rejected an image on quality grounds. E.g., one person said “*I would probably look for a real image first. But, if I couldn't find one then I guess this one would probably be good enough*”.
- **Potential actions to reduce misguided rejections:**
 - Provide good examples during onboarding and always accessible help guidance on the tool.
 - Ask for feedback when an image suggestion is rejected (through a simple set of radio button responses). Prompt users to reevaluate if they select certain reasons. For example, if the reason is “not related enough”, show a hint that a detailed caption can be written to connect the image more directly to the article.

Adding a caption

Task summary

After reviewing at least 20 images, participants were asked to look in detail at the “Add caption” dialog shown after selecting “Yes” on a suggested image. They were asked to explain their understanding of what this caption was for and where it went, and then write a caption for one image based on this understanding.

Prior to this specific captions task, almost all participants wrote captions that were more like **justifications** for adding the image to the article, seemingly assuming this was an extension of the image matching task.



Adding a caption

Sample captions from the “Add caption” task.

When participants were asked to look more closely at the add captions dialog, some did provide better captions that were less justification-heavy.

Even so, most captions were still not relevant, contained inaccurate assumptions, were written in a redundant format, and/or was too informal in tone.

5 Good captions:

- “Maximiliano de Almeida flag” (Image 16)
- “Image of the Jocara snout moth.” (Image20)
- “Jocara snout moth” (Image 20)
- “A type of Jocara moth.” (Image 20)
- “Copy of Islendingabok.” (Image 35)

4 Fair captions (format, tone, or embellishment):

- “A picture of the author at the Montreal Book Festival, taken by a credible source.” (Image 15) - “credible source” is embellished
- “This is the logo for ash-shiraa's magazine.” (Image 18) - formatting.
- “The 1894 baseball event” (image 30) - incorrect information unrelated to image.
- “This is a picture of Janusz Kruk in 1972” (Image38) - format and some details are inaccurate (it is not from 1972)

6 Bad captions (not describing image, or giving wrong info):

- “The article is talking about Nora, the image is Nora with her ski gear” (Image 14)
- “Good picture for the bug” (Image 20)
- “Sante Fe Class 1158, locomotive, trains, trains sketch” (Image 23) - caption is written as a series of tags meant for SEO
- “This is a crystal ball from the German fairy tale” (Image 42) - it’s an unrelated crystal ball photo.
- “Matched well I understand great lakes better” (Image 58) - justification and non-neutral tone
- “The entrance to the commune” (Image 47) - user incorrectly assumes this is the entrance.

Adding a caption

Detailed findings and observations

- **Understanding of where captions appeared and their purpose was low.** Wrong answers included:
 - Confusing captions with SEO tags
 - Confusing it with the Commons image caption
 - Confusing it with the Suggestion reason
 - Thinking it would appear in the “Edit history” (at least two people gave this rationale)
- **A justification reason for using the image was written by the majority of participants,** instead of describing the image with reference to the article. One person (*thetidebreaks*) specifically mentioned the hint text (“Briefly describe how this image relates to the article”) as making her think that was the intent.
- **Image descriptions and filename was usually used on the more accurate captions.** A few testers sought to copy and paste directly from the description field.
- **Quality of captions written was not good.**
 - **Redundant text** – e.g., “*Labeling says this is a house in the village*”
 - **Informal language** – “*It’s a picture of him and his family*”
 - **Not neutral** – e.g., “*Statue of a powerful Roman figure*”
 - **Embellishments/Assumptions added** – e.g., “*A picture of the author at the Montreal Book Festival, taken by a credible source.*”
 - **Overly descriptive (akin to alt-text)** – e.g., “*It shows the land layout on a map. Shows where the territory begins and ends. Also shows the divisions.*”
- **Potential actions to improve caption-writing:**
 - Update the hint and label to clarify the intent
 - Provide examples of good captions
 - Show how the image and caption will appear in the context of the article.
 - Show the Commons description or caption (or filename if neither are available) as a starting placeholder (or visible) during caption-adding.
-

Article and image info

Summary

Participants rated the usefulness of each of the different pieces of information shown in helping them to evaluate the suggestions towards the end of the test.

Each item was rated from 1 to 5:

- 1 = Not useful at all
- 5 = Very useful

1 2 3 4 5

○ ○ ○ ○ ○

< Not at all useful Very useful >



Suggestion reason

Another user thinks it's related (image is in a Commons category linked to the article's Wikidata item)

Filename

Oprtalj - Crkva Sveti Roka - 01.jpg

3456 x 2304 pixels

03 Sep 2013

Maesi64

[More info \(Commons file page\)](#)

Image description

English: Church of Saint Roch (Crkva Sv. Roka/ Chiesa di San Rocco), in Opatalj, Croatia

Image caption

-

Categories

Saint Roch church (Opatalj), Saint Roch churches, 1470s churches in Croatia, Built in Croatia in 1581, Churches built in 1581

Tags

-

Is this a good image for the following article?

[Church of Saint Roch](#)

Church of Saint Roch or Saint Roch's Church may refer to:

Yes

No

Unsure

Suggestion reason – this is the source of the algorithm's recommendation.

There are three possible reasons:

1. Another user thinks this is a strong match (used in the Wikidata item)
2. Another user thinks it's related (image is in a Commons category linked to the articles Wikidata item)
3. Used in the same article in another language/project <list of wikis>

Image metadata:

Filename – image name + file type

Size – image size in pixels

Date – creation date

Username – of the author

More info – link to its Commons file page

Image description – file page description

Image caption – file page caption

Categories – Commons file page categories

Tags – Structured data depicts tags

Article information:

Article title with link to full article

Article extract – Beginning sentences of the article.

Article and image info

Aggregate ratings

Overall, participants felt the information presented was sufficient for their decision-making process and didn't feel additional types of information were missing. However, they did want highly-rated items such as image descriptions and image captions to be more complete (some images were missing either or both fields).

Notes:

- Some ratings were affected by participants being confused by two pieces of information (e.g., confusing Tags with Categories)
- People sometimes evaluated based on info that is good for the image (e.g. Date) rather than their decision-making process.

Rating 1 (Not useful) to 5 (Very useful)

Information	Mean	Median	Spread
Suggestion reason	3.7	4	
<i>Image info</i>			
Filename	3.7	4	
Size	2.5	2	
Date	3.3	3	
Username	2.1	2	
Link to More info (Commons file page)	3.4	4	
Description (when available)	4.5	5	
Caption (when available)	4.3	4	
Categories	3.7	4	
Tags (when available)	2.9	3	
<i>Article info</i>			
Link to the full article	4.1	5	
Article extract	3.9	4	

Article and image info

1 of 2 – Detailed findings and observations on how participants' used with the available data points

- Suggestion reason was only a clear and strong indicator when the reason was the image has been used in one or more languages/projects.
- Other suggestion reasons were poorly understood and ignored. That is, when the reason was either related to the Wikidata item (“Another user thinks this is a strong match...”) or Categories “Another use thinks it’s related...”), participants wanted to have more information about who that other user was and their reasons for the match.
 - **Action:** Potentially revise info presentation to separately allow users to open the other language/project pages when this is shown, and decouple this with the other ‘reasons’.
 - **Action:** Revise copy about the other reasons to remove a specific “Another user”, and provide more explanation about wikidata and categories
- Image descriptions and captions (when available) were rated the highest overall.
- Link to read more of the article was also highly rated despite being rarely used. This may have been related to participants wary of leaving the prototype in the test conditions.
- **Article extracts were important but often not long enough**, according to many participants.
 - **Action:** Potentially mitigated by making a more overt action to view the entire article
 - **Action:** Explore a layout that allows easy access to a longer article preview
- **A couple of users misunderstood article extract to be part of the image metadata.** This again led to incorrect, false positives.

Article and image info

2 of 2 – Detailed findings and observations on how participants' used with the available data points

- **Date was deemed very important by a few testers for certain article topics (e.g., events).** It was also a factor in assessing recency/freshness for other article subjects (biographies).
- **“Technical” labels like image size and Tags were less understood and prone to misinterpretation.**
 - Image size – was misunderstood to refer to how it appears on the screen layout (“all the pictures come up the same size in your iPhones”) or file size (“size limit on the wiki?”).
 - Tags – were misunderstood to be Categories or SEO keywords. (“[The tags] were medium useful because that's just a person's opinion”)
- **Confusion regarding very similarly named data types and labels.** For example Tags vs Categories, the Description vs caption, the SD caption on Commons vs the caption for the image on the article.
- **Lower placement of metadata** led some testers to not see or give importance to the information.
 - **Action:** Consider changing positions of certain information to arrange most to least important
 - **Action:** Do not show anything for blank fields
- **Some participants suggested not showing information deemed not useful,** with “Username” being rated lowest overall. No one actually mentioned using this piece of information in their deliberations.
 - **Action:** Consider removing this Username.
- **Google is mentioned as a way to find out more info/verification when “further research” is needed.** This may be related to a lack of understanding around image provenance.
 - **Action:** Provide information relating to the source and reliability of image metadata

Usability and interaction behaviour

- **Areas of improvement/other exploration observed**
 - Layout explorations that draw focus to article context for the image suggestion
 - Increase affordance that the whole article can be accessed
 - Increase affordance of the zoom function
 - Increase affordance to viewing more information on the Commons page
 - Provide clear pathways back from Commons file pages and the full article
 - Increase proximity of navigation buttons from edges of mobile screens – On some devices, the “Next” and “Prev” buttons being so close to the bottom interfered with native navigation.
- **Specific suggestions from participants**
 - Providing more than one image suggestion for an article, especially in reference to after an image is rejected for an article.
 - Asking users to select/enter a reason for their “No” rejection.
- **Some things that worked well**
 - Copy changes for Group C that made the clearer connection between article and image being matched for 4 of 5 testers in Group C.
 - Having the ability to zoom and pan the image
 - Showing all the metadata below the image seemed to be clear for users. This was true even for metadata shown with labels only (size, date, author)

General observations

Other notes (comments about perceptions of Wikipedia, images, etc)

- **People are interested in adding images this way.**
Everyone said they would be interested in doing this task, some citing they wanted “give back” to Wikipedia, while others just found this activity “fun”, “easy” and “relaxing” to do.
- **Relatively low recognition that Wikipedia can be edited by anyone** - 5 of 15 participants didn't know it could be edited prior to the test. Notably, these participants were no less able to complete the task.
- **Zero recognition of Wikimedia Commons project** – no one mentioned knowing Commons was where the images were from, despite three people saying they had used Wikimedia Commons in the screener.
- **Google confusion** - users thought that the suggestions came from Google images and Google search results. This may be related to the low recognition of Commons, and limited explanation in the prototype.
- **Workflow and context assumptions** by a few testers^[1] that this would appear for a person writing the article, being offered suggestions for an image to add during article creation. This makes them think that the actual reviewer would know more about the subject to make the decision.

[1] Participant “Nutrition” at ~2m13s, and participant “jsts” at ~6m20s.

Choice quotes

“And the information is very clear here that it says it's– the same image was used in the same article in a different language article. So it's properly added to the article, so I will click here yes.”

– [jsts \(good use of the “Suggestion reason”\)](#)

“Well I mean, yes, it's an archaeological site. I don't really see the relevance [to the article 177BC]. It's not like Wikipedia is full of random pictures.”

– [SerendipityTester \(reviewing with reference to Wikipedia\)](#)

“...because the language in the image does not correspond to, I guess it's English? It's not a good image to use because people will not understand”

– [Lex b \(importance of language of images and metadata\)](#)

“Yeah, because I need some context as to whether or not this– the images– you know, what image would be the right context for the article.”

– [SilverGirl47 \(importance of access to the full article\)](#)

““I mean I can frickin' do this all day, let me tell ya – I love finding out things like this!””

– [thetidebreaks \(Interest in the task\)](#)

“...if the article is about Ken Richardson, I'd probably want to pick a picture on him. But if I could use two pictures, then I would like to use the racing car as well, because it is a nice visual and it would add to the article”

– [Nutrition \(single main vs multiple images\)](#)

Choice quotes

"I don't like that photo. It's fuzzy and it's—I don't know if it's accurate."

- [cpmcbusertester \(image quality mixed with suggestion quality\)](#)

"Just having a—like, one person of color next to the white girl with dreadlocks isn't very diverse."

- [Lady Burzum \(critical thinking on suitability\)](#)

"I don't think there really needs to be a lot of description, but I mean, it would just help clarify what they were looking at, and why the picture was part of the article"

- [SerendipityTester \(caption understanding\)](#)

"...at first I didn't really understand what was the image about but then I found the article at the bottom and everything made sense. So after that everything became smooth"

- [Lex_b \(Written answer about Article card issues\)](#)

"Maybe there's almost too much information to scroll through. It needs to be organized in a sense. But it's pretty good"

- [Nutrition \(image info\)](#)

"So what I'm seeing is the picture of the bird, information about that picture, and that bird. And then on the bottom it's asking me if I would like to and that image to the article. So I'm thinking the person who writes the article and then looks for an image and this is one of the images that would come up."

- [Nutrition \(thinking tool is used during article creation, also that multiple choices are available\)](#)

Overall recommendations & next steps

Recommendations

- Expand the test to a much larger sample size to verify and correct the quant-based initial findings
- Include non-English speaking participants in other language wikis in further tests
- Explore the potential for variants relating for Android MVP, including but not limited to:
 - Revising information hierarchy to emphasise article and reducing image metadata to key information only
 - Different ways to capture captions
 - Incorporating rejection reason feedback
- Explore and apply relevant proposed usability and interaction recommendations in the Android MVP

Next steps

- Share findings with Android team and Growth team
- Share findings with SD & Design strategy team members working on Media Matching research
- Share findings with wider Image recommendations working group
- Conduct another round of testing in Arabic Wikipedia with new sample data(?)