

Mind – Formative Evaluation

**Daisy – Eliciting Richer  
Conversations at a Chatbot**

Joyce Ma

October 2007

THIS IS NOT A DEFINITIVE FINAL REPORT

FORMATIVE evaluation studies like this one often:

- **are conducted quickly**, which may mean
  - small sample sizes
  - expedited analyses
  - brief reports
  
- **look at an earlier version** of the exhibit/program, which may mean
  - a focus on problems and solutions, rather than successes
  - a change in form or title of the final exhibit/program

# Mind – Formative Evaluation

## Daisy – Eliciting Richer Conversations

Joyce Ma  
October 2007

PURPOSE .....	2
KEY FINDINGS.....	3
STUDY 1.....	5
Method .....	5
Data Collected.....	7
Results .....	8
Which questions prompted visitors to respond? .....	8
Which question formats engaged visitors in conversation? .....	11
STUDY 2.....	12
Method .....	12
Data Collected.....	13
Results .....	14
Is a 2-Part question better than a Complex question in encouraging visitors to give a richer, more elaborate response? .....	14
Does the second part of a 2-part question elicit more elaborated answers?.....	15
For a 2-Part question, is a more general follow-up question better than a specific follow-up question? .....	18
Is a question that asks directly about the visitor's thinking better at fostering self-reflection than one that asks about machine thinking? .....	19
RECOMMENDATIONS.....	20
ACKNOWLEDGEMENTS .....	21

### PURPOSE

Daisy (Figure 1) is a chatbot that tries to engage visitors in conversation. As part of its program, Daisy asks canned questions, intended to have the visitor reflect on the nature of his/her own thinking and feeling as well as those of the machine.

Figure 1. Daisy



This report documents two formative evaluation studies conducted on Daisy to determine:

- Which questions encouraged visitors to reflect on their own and/or Daisy's thinking and feeling? These evaluations served to identify a set of canned questions to be included in the Daisy exhibit.
- What question formats engaged visitors in conversation? As well as specifically informing Daisy, we hope that the findings will find application to a broader class of multimedia exhibits that engage visitors in dialog for learning.

## KEY FINDINGS

The following summarizes the key findings from both evaluations.

- Questions that ask directly about thoughts and emotions are better at eliciting conversation on those topics. Along those same lines, questions about visitors themselves are better at prompting self-reflection. For example,

DAISY > How do I know I'm talking to a human and not just another machine?

tends to promote more self-reflection than

DAISY > Are you sure that I'm not a real person talking to you by e-mail?

DAISY > What would it take to convince you?

- Format matters. Different question formats (i.e., Simple question, Complex question, 2-Part question) elicited different types of answers: no answer, simple answer, or elaborated answer. More specifically,
  - A 2-Part format (i.e. a Simple question with a yes/no answer followed by a question that asks for elaboration) is comparable if not better than its Complex counterpart (a question that asks for more than a yes/no response) in eliciting richer answers. For instance, a higher percentage of visitors gave elaborated responses to

DAISY> Are you usually a logical person?

DAISY> Give me an example.

compared to

DAISY> Are you usually a logical person, or do you let your feelings affect your decisions? Give an example of a recent logical or emotional decision you made.

- A follow-up question can prompt for elaboration from visitors.
- Visitors were more likely to give an elaborated answer when a more general prompt was used as the follow-up question. For example,

DAISY > Are there some stereotypes that you believe?

DAISY > Why do you think that?

was more effective than

DAISY > Are there some stereotypes that you believe?

DAISY > Why do you think that some stereotypes are true?

## STUDY 1

In this study we looked at 7 canned questions. One fires every 8th time that Daisy ‘speaks’. To help us gauge which format may be better at eliciting richer conversations, we included different types of question formats as shown in Table 1.

**Table 1. Tally of types of answers given for different format of questions.**

Question Format	Canned Question
<b>Simple question</b> - questions that can be answered with a one-word answer.	DAISY > Are you sure that I'm not a real person talking to you by e-mail?
	DAISY > Are there some stereotypes that you believe?
	DAISY > Are you having a conversation with me, or with yourself?
<b>Complex question</b> - questions that ask for elaborations, that cannot be answered with a one-word answer.	DAISY > Does the fact that my name is Daisy affect the way we interact? What if my name were Ludwig or Prudence instead?
	DAISY > My mind works by recombining and regurgitating things that people tell me. But how is that different than what your mind does?
<b>2-Part question</b> - A simple question fires first. Daisy follows the visitor's immediate response with a complex question.	DAISY > Are you usually a logical person?
	DAISY > Give me an example
	DAISY > If I say something that makes sense, does that mean I was thinking? DAISY > If I wasn't thinking, where did the thought come from?

### Method

We logged visitors’ conversations with Daisy and extracted the line immediately following a canned question, and then coded that line according to whether the visitor gave

- No answer. These include remarks that do not in any way answer Daisy’s question, carriage returns, and function key presses. The following are different examples of visitor input that fell into the ‘No Answer’ category:

```
DAISY> My mind works by recombining and regurgitating things
that people tell me. But how is that different than what
your mind does?
```

```
HELLO> i dontcare
```

```
DAISY> Are you sure that I'm not a real person talking to you
by e-mail?
```

```
JUSTIN>
```

```
DAISY> Are there some stereotypes that you believe?
```

```
EMILY> %%%f1%%%
```

- **Simple answer.** These are one-word answers or equivalent to Daisy's questions. For instance,

DAISY> Are there some stereotypes that you believe?

JADE> no

DAISY> Does the fact that my name is Daisy affect the way we interact? What if my name were Ludwig or Prudence instead?

JENNY> it would not affect the way we interact.

DAISY> Are you having a conversation with me, or with yourself?

JENNY> i am conversing with you.

- **Elaborated answers.** These are answers to Daisy's questions that go beyond a simple answer. They may include an explanation of a simple answer or comments that otherwise advance the conversation in response to that particular question from Daisy. For example,

DAISY> My mind works by recombining and regurgitating things that people tell me. But how is that different than what your mind does?

HANS> touche' but still. you have no idea of what you are saying. there is no "you." you simply eam of code

DAISY> Are you having a conversation with me, or with yourself?

DOM> techniquely wwith myself because you are dumb as heck...

DAISY> Are there some stereotypes that you believe?

KAT> .....what the heck is a sterotype? i've heard of it, but i have no clue what thais!!!!

We also looked at each elaborated answer to see if it includes comments about:

- The visitor's own (or human) thinking and/or feeling, or **self-reflection**, such as

DAISY> Does the fact that my name is Daisy affect the way we interact? What if my name were Ludwig or Prudence instead?

MR. MUSHROOM> if your name was ludwig or prudence i would like you more

DAISY> My mind works by recombining and regurgitating things that people tell me. But how is that different than what your mind does?

SOFIA> cause i think by maself and nobody helps me u no

DAISY> Are you usually a logical person?

LULU> yes

DAISY> Give me an example.

LULU> i think to much of something before i do it ... like ... talking to a boy...

- Daisy's (or machine) thinking and/or feeling. For example

DAISY> Are you sure that I'm not a real person talking to you by e-mail?

MAX> yes, because your talking like a retard

DAISY> My mind works by recombining and regurgitating things that people tell me. But how is that different than what your mind does?

MARILYN> are you like hal

DAISY> If I say something that makes sense, does that mean I was thinking?

SCOTT> maybe

DAISY> If I wasn't thinking, where did the thought come from?

SCOTT> there was no thought, only words

### **Data Collected**

We collected computer logs from 6/22/2007 to 7/19/2007, giving us a total of 2484 conversations<sup>1</sup>. Table 2 shows the number of times each of our canned questions fired in these conversations.

**Table 2. Tally of the number of times each canned question appears**

Question Format	Canned Question	Total times fired
Simple	DAISY > Are you sure that I'm not a real person talking to you by e-mail?	123
Simple	DAISY > Are there some stereotypes that you believe?	119
Simple	DAISY > Are you having a conversation with me, or with yourself?	118
Complex	DAISY > Does the fact that my name is Daisy affect the way we interact? What if my name were Ludwig or Prudence instead?	121
Complex	DAISY > My mind works by recombining and regurgitating things that people tell me. But how is that different than what your mind does?	119
Part 1 of 2 Part	DAISY > Are you usually a logical person? (part 1 only)	125
Part 2 of 2 Part	DAISY > Give an example (part 2 only)	122
Part 1 of 2 Part	DAISY > If I say something that makes sense, does that mean I was thinking? (part 1 only)	121
Part 2 of 2 Part	DAISY > If I wasn't thinking, where did the thought come from? (part 2 only)	119

<sup>1</sup> A new conversation is created when a visitor enters a username to begin his/her chat with Daisy.

## Results

### Which questions prompted visitors to respond?

Table 3 tallies the types of answers visitors gave for each canned question.

**Table 3. Tally of types of answers given for each question.** Counts at or above 50% are shaded orange.

Format	Canned Question	No Answer	Simple Answer	Elaborated Answer	Total times fired
Simple	DAISY > Are you sure that I'm not a real person talking to you by e-mail?	42 (34%)	62 (50%)	19 (15%)	123
	DAISY > Are there some stereotypes that you believe?	51 (43%)	47 (39%)	21 (18%)	119
	DAISY > Are you having a conversation with me, or with yourself?	30 (25%)	71 (60%)	17 (14%)	118
Complex	DAISY > Does the fact that my name is Daisy affect the way we interact? What if my name were Ludwig or Prudence instead?	42 (35%)	46 (38%)	33 (27%)	121
	DAISY > My mind works by recombining and regurgitating things that people tell me. But how is that different than what your mind does?	47 (39%)	23 (19%)	49 (41%)	119
2-Part	DAISY > Are you usually a logical person? Give me an example (Taken as a whole <sup>2</sup> )	31 (25%)	25 (20%)	68 (55%)	124
	DAISY > Are you usually a logical person? (part 1 only)	35 (28%)	72 (58%)	17 (14%)	124
	DAISY > Give an example (part 2 only)	50 (40%)	12 (10%)	62 (50%)	124
	DAISY > If I say something that makes sense, does that mean I was thinking? If I wasn't thinking, where did the thought come from? (Taken as a whole <sup>1</sup> )	31 (26%)	22 (18%)	66 (55%)	119
	DAISY > If I say something that makes sense, does that mean I was thinking? (part 1 only)	39 (32%)	67 (55%)	15 (12%)	121
	DAISY > If I wasn't thinking, where did the thought come from? (part 2 only)	38 (32%)	17 (14%)	64 (54%)	119

<sup>2</sup> If the visitor gave an elaborated answer for either part of the 2-part question, then the answer, taken as a whole, was coded as an elaborated answer. Alternatively, if the visitor gave a simple answer and NOT an elaborated answer for either part, then that answer counted as a simple answer.

Table 4 shows the types of elaborated answers visitors gave to Daisy's canned questions. In addition to coding for responses that comment on their own or Daisy's thinking and feeling, we looked to see if visitors asked Daisy for help in answering. For example,

DAISY> Are there some stereotypes that you believe?

ZACK> whgat is a stereo type?

**Table 4. Tally of types of elaborated answers according to question.** Responses that contain comments about their own or Daisy's thinking and feeling, are coded as both.

Format	Canned Question	Self-Reflection	Daisy	Help	Other <sup>3</sup>	Total Elaborated Answers
Simple	DAISY > Are you sure that I'm not a real person talking to you by e-mail?	6 (32%)	14 (74%)	0 (0%)	1 (5%)	19
	DAISY > Are there some stereotypes that you believe?	7 (33%)	5 (24%)	8 (24%)	5 (24%)	21
	DAISY > Are you having a conversation with me, or with yourself?	7 (41%)	11 (65%)	0 (0%)	5 (30%)	17
Complex	DAISY > Does the fact that my name is Daisy affect the way we interact? What if my name were Ludwig or Prudence instead?	14 (42%)	5 (15%)	0 (0%)	14 (42%)	33
	DAISY > My mind works by recombining and regurgitating things that people tell me. But how is that different than what your mind does?	40 (82%)	15 (31%)	0 (0%)	0 (0%)	49
2-Part	DAISY > Are you usually a logical person? Give me an example (Taken as a whole <sup>4</sup> )	42 (62%)	3 (4%)	7 (10%)	21 (31%)	68
	DAISY > Are you usually a logical person? (part 1 only)	12 (71%)	1 (6%)	0 (0%)	4 (24%)	17
	DAISY > Give an example (part 2 only)	35 (56%)	2 (3%)	7 (11%)	18 (29%)	62
	DAISY > If I say something that makes sense, does that mean I was thinking? If I wasn't thinking, where did the thought come from? (Taken as a whole <sup>4</sup> )	6 (9%)	45 (68%)	0 (0%)	24 (36%)	66
	DAISY > If I say something that makes sense, does that mean I was thinking? (part 1 only)	3 (20%)	10 (67%)	0 (0%)	0 (0%)	15
	DAISY > If I wasn't thinking, where did the thought come from? (part 2 only)	4 (6%)	41 (64%)	0 (0%)	20 (31%)	64

<sup>3</sup> This category includes responses that did not comment on the visitor's own or Daisy's thoughts or emotions. For instance,

DAISY> Does the fact that my name is Daisy affect the way we interact? What if my name were Ludwig or Prudence instead?

CONNOR> all those names are stupid

<sup>4</sup> If the visitor gave an elaborated answer for either part of the 2-Part question, then the answer, taken as a whole, was coded as an elaborated answer. Alternatively, if the visitor gave a simple answer and NOT an elaborated answer for either part, then that response counted as a simple answer.

The above tables suggest that *promoting self-reflection depends on the nature of the question asked*. More specifically,

- The results indicate that the following questions work well in eliciting elaborated answers about visitors' own thinking and feeling:

DAISY> My mind works by recombining and regurgitating things that people tell me. But how is that different than what your mind does?

DAISY> Are you usually a logical person? Give me an example

- Alternatively, some questions, though good at eliciting elaborated answers, tended to focus visitors on Daisy's rather than their own thinking and feeling

DAISY> If I say something that makes sense, does that mean I was thinking? If I wasn't thinking, where did the thought come from?

- Other questions, elicited more answers focused on topics other than thought and emotions. More than one-third of the elaborated answers to

DAISY> Does the fact that my name is Daisy affect the way we interact? What if my name were Ludwig or Prudence instead?

referred to the names themselves.

- And, some questions were more confusing than others:

DAISY> Are there some stereotypes that you believe?

DAISY> Are you usually a logical person? Give me an example

About one-quarter of visitors who tried to give an elaborated response asked for help with these question; many of these visitors simply did not know what *stereotype* means.

Likewise, a few visitors did not knowing what would count as an example.

### Which question formats engaged visitors in conversation?

To see if different question *formats* gave us different types of answers, we tabulated the results according to question format. (See Table 1 and Table 5).

**Table 5. Tally of types of answers given for different formats of questions.**

Question Format	No Answer	Simple Answer	Elaborated Answer	Total
Simple question	123 (34%)	180 (50%)	57 (16%)	360
Complex question	89 (37%)	69 (29%)	82 (34%)	240
2-Part question	30 (11%)	95 (35%)	145 (54%)	243

This rough analysis suggests that *different formats tend to elicit different types of answers*. The 2-Part question received more elaborated answers, while simple questions tend to get simple answers. We will take a more detailed look at the types of responses different formats tend to encourage in Study 2.

## STUDY 2

In this second evaluation study, we took a closer look at the question format for some of the more successful canned questions we tested in the previous study. More specifically, this evaluation determined:

- Was a 2-Part question better than a Complex question in encouraging visitors to give a richer, more elaborate response?
- Did the second part of a 2-part question elicit more elaborate responses?
- For a 2-Part question, was a more general follow-up question better than a specific follow-up question?
- Was a question that asks directly about the visitor's thinking better at fostering self-reflection than one that asks about machine thinking?

We designed a set of questions as listed in Table 1. Visitors' responses to these and the set of canned questions from Study 1 were used to answer the above questions.

**Table 6. Canned questions used in Study 2.** Each question fires every 8<sup>th</sup> time Daisy 'speaks' during a conversation.

Question Format	Canned Question
<b>Complex question</b> - questions that ask for elaborations, that cannot be answered with a one-word answer.	DAISY > How do I know I'm talking to a human and not just another machine?
	DAISY > Are you usually a logical person, or do you let your feelings affect your decisions? Give an example of a recent logical or emotional decision you made.
<b>2-Part question</b> - A simple question fires first. Daisy follows the visitor's immediate response with a complex question.	DAISY > My mind works by recombining and regurgitating things that people tell me. Doesn't your mind do the same thing? In what ways are our minds different?
	DAISY > Are you sure that I'm not a real person talking to you by e-mail?
	DAISY > What would it take to convince you?
	DAISY > Are there some stereotypes that you believe? DAISY > Why do you think that?
	DAISY > Are there some stereotypes that you believe? DAISY > Why do you think that some stereotypes are true?

### Method

We followed the same procedure we used in Study 1. That is, we collected the computer log of visitors' conversations with Daisy. Then, we extracted the line immediately following a canned question from the log, and, using the same coding scheme as before, coded that line according to whether the visitor gave:

- No Answer

- A Simple Answer
- An Elaborated Answer
  - A Self-Reflective Answer
  - An Answer about Daisy
  - Answers Other than about the self or Daisy

### **Data Collected**

We logged 2219 conversations between 9/5/2007 and 10/9/2007. Table 2 tallies the number of times each of our canned questions fired.

**Table 7. Tally of the number of times each canned question appears**

<b>Question Format</b>	<b>Canned Question</b>	<b>Total times fired</b>
Complex	DAISY > How do I know I'm talking to a human and not just another machine?	137
Complex	DAISY > Are you usually a logical person, or do you let your feelings affect your decisions? Give an example of a recent logical or emotional decision you made.	141
Part 1 of 2 Part	DAISY > My mind works by recombining and regurgitating things that people tell me. Doesn't your mind do the same thing?	143
Part 2 of 2 Part	DAISY > In what ways are our minds different?	138
Part 1 of 2 Part	DAISY > Are you sure that I'm not a real person talking to you by e-mail?	145
Part 2 of 2 Part	DAISY > What would it take to convince you?	142
Part 1 of 2 Part	DAISY > Are there some stereotypes that you believe?	128
Part 2 of 2 Part	DAISY > Why do you think that?	127
Part 1 of 2 Part	DAISY > Are there some stereotypes that you believe?	155
Part 2 of 2 Part	DAISY > Why do you think that some stereotypes are true?	138

## Results

Is a 2-Part question better than a Complex question in encouraging visitors to give a richer, more elaborate response?

To answer this question, we compared two sets of Complex and 2-Part questions.

	Set 1	Set 2
<b>Complex Question</b>	DAISY> Are you usually a logical person, or do you let your feelings affect your decisions? Give an example of a recent logical or emotional decision you made.	DAISY> My mind works by recombining and regurgitating things that people tell me. But how is that different than what your mind does?
<b>2-Part Question</b>	DAISY> Are you usually a logical person? DAISY> Give me an example.	DAISY> My mind works by recombining and regurgitating things that people tell me. Doesn't your mind do the same thing?  DAISY> In what ways are our minds different?

For Set 1, we found that visitors answered differently for the different question formats,  $\chi^2(2, N = 283) = 35.1, p = .000$ . Furthermore, when we compared the number of elaborated answers according to the type of question format, we found that visitors were more likely to give elaborated answers to a 2-Part question; Fisher's Exact Test,  $p < 0.0001$ . See Table 8.

**Table 8. Comparison of Complex versus 2-Part Format (Set 1)– Tally of types of answers according to type of question format**

	Set 1	No Elaborated Answer		Elaborated Answer	Total
		No Answer	Simple Answer		
<b>Complex Question</b>	DAISY> Are you usually a logical person, or do you let your feelings affect your decisions? Give an example of a recent logical or emotional decision you made.	85 (60%)	21 (15%)	35 (25%)	141
<b>2-Part Question</b>	DAISY> Are you usually a logical person? DAISY> Give me an example. (Taken as a whole <sup>5</sup> )	31 (25%)	25 (20%)	68 (55%)	142

On the other hand, for Set 2, we did not detect a statistically significant difference in the percentage of visitors who answered differently for the Complex versus the 2-Part question

<sup>5</sup> If the visitor gave an elaborated answer for either part of the 2-part question, then the answer, taken as a whole, was coded as an elaborated answer. Alternatively, if the visitor gave a simple answer and NOT an elaborated answer for either part, then that answer counted as a simple answer.

format,  $\chi^2(2, N = 257) = 3.21, p = .201$ . See Table 9. However, there is a larger, but statistically insignificant (Fisher's Exact Test) number of visitors who gave any answer, simple or elaborated, to those who did not, between the two formats.

**Table 9. Comparison of Complex versus 2-Part Format (Set 2)– Tally of types of answers according to type of question format**

	Set 2	No Answer	Any Answer		Total
			Simple Answer	Elaborated Answer	
<b>Complex Question</b>	DAISY> My mind works by recombining and regurgitating things that people tell me. But how is that different than what your mind does?	47 (39%)	23 (19%)	49 (41%)	119
<b>2-Part Question</b>	DAISY> My mind works by recombining and regurgitating things that people tell me. Doesn't your mind do the same thing? DAISY> In what ways are our minds different? (Taken as a whole)	40 (29%)	33 (24%)	65 (47%)	138

The above suggests that *a 2-Part format is comparable if not better than its Complex counterpart in eliciting richer responses.*

#### Does the second part of a 2-part question elicit more elaborated answers?

Figure 2, Figure 3, Figure 4, and Figure 5 show the number of visitors who responded to part 1 and part 2 of our 2-Part questions. Notice that in all cases, if a visitor did not respond to part 1, they were not likely to give an elaborated answer to part 2; only 1% to 6% did so. However, 14% to 26% of the visitors who gave a simple answer to part 1 of a 2-Part question did go on to elaborate on their answers in part 2 when prompted. Although not a high percentage, this finding does indicate that *adding a follow-up question can prompt for additional elaboration from visitors.*

Figure 2. Treemap for “My mind works by recombining and regurgitating things that people tell me. Doesn't your mind do the same thing? In what ways are our minds different?”

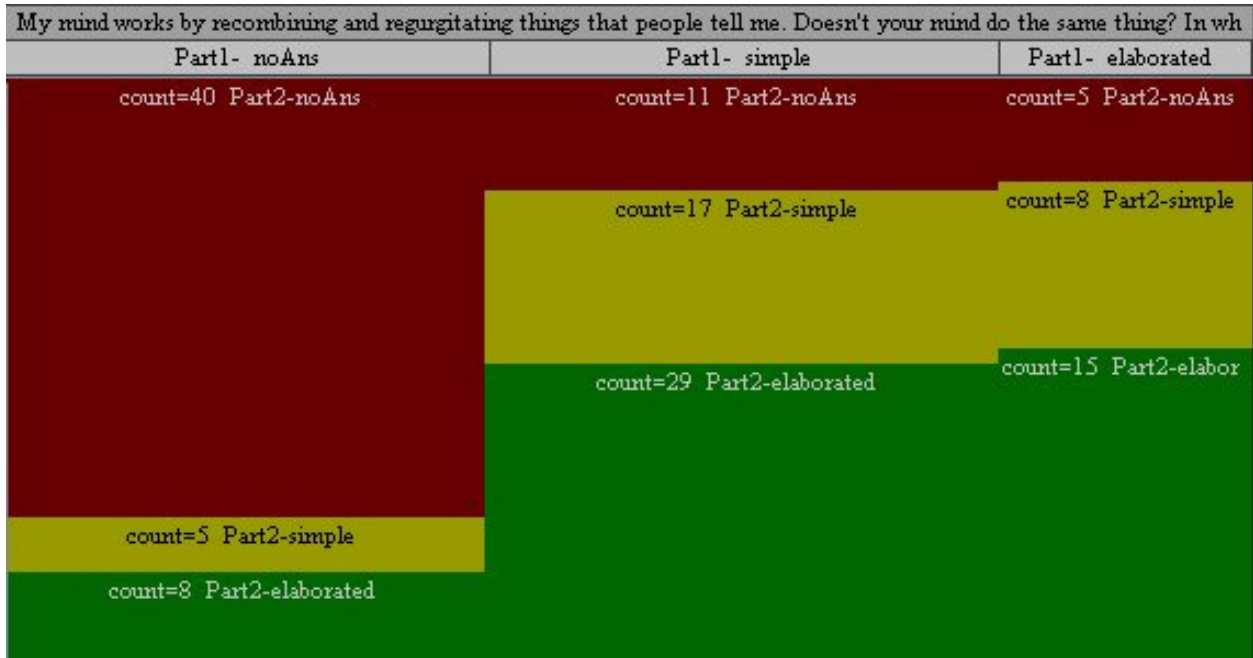


Figure 3. Treemap for “Are you sure that I'm not a real person talking to you by e-mail? What would it take to convince you?”

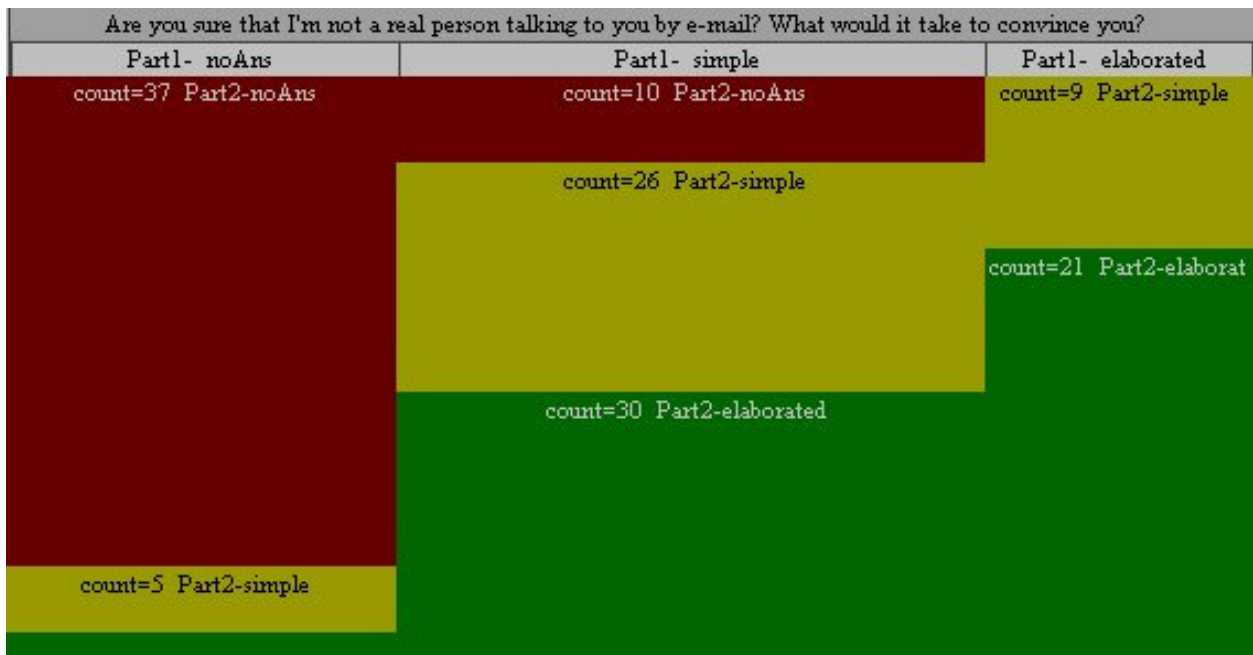


Figure 4. Treemap for “Are there some stereotypes that you believe? Why do you think that?”

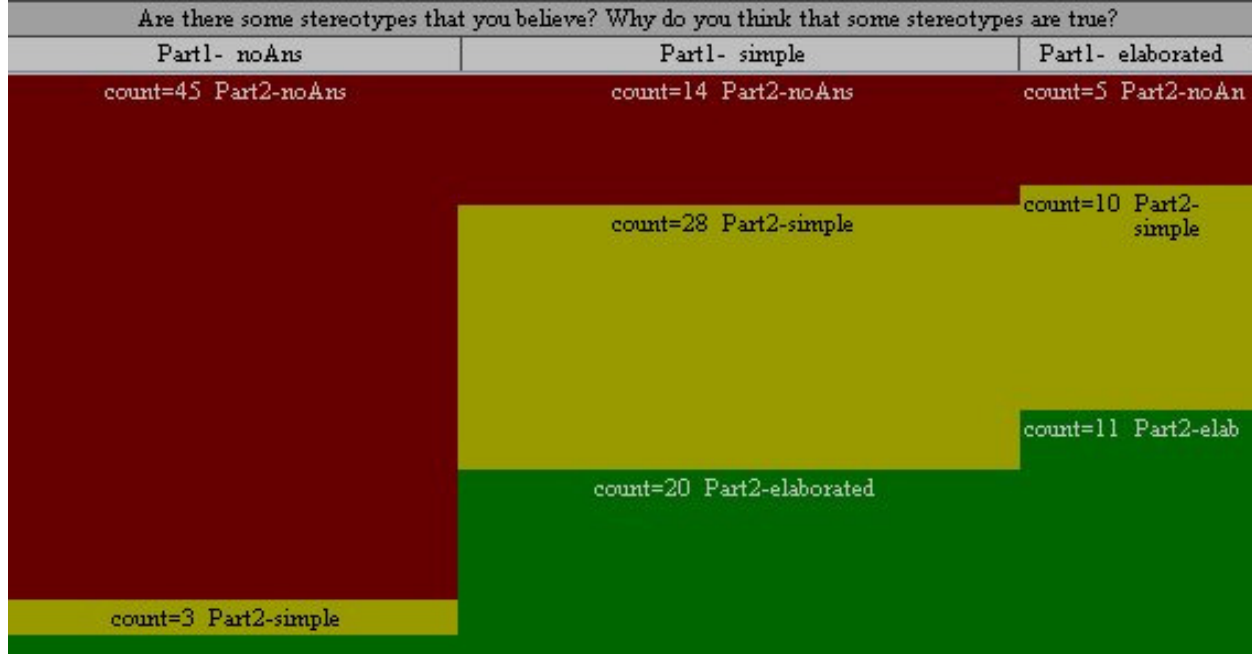
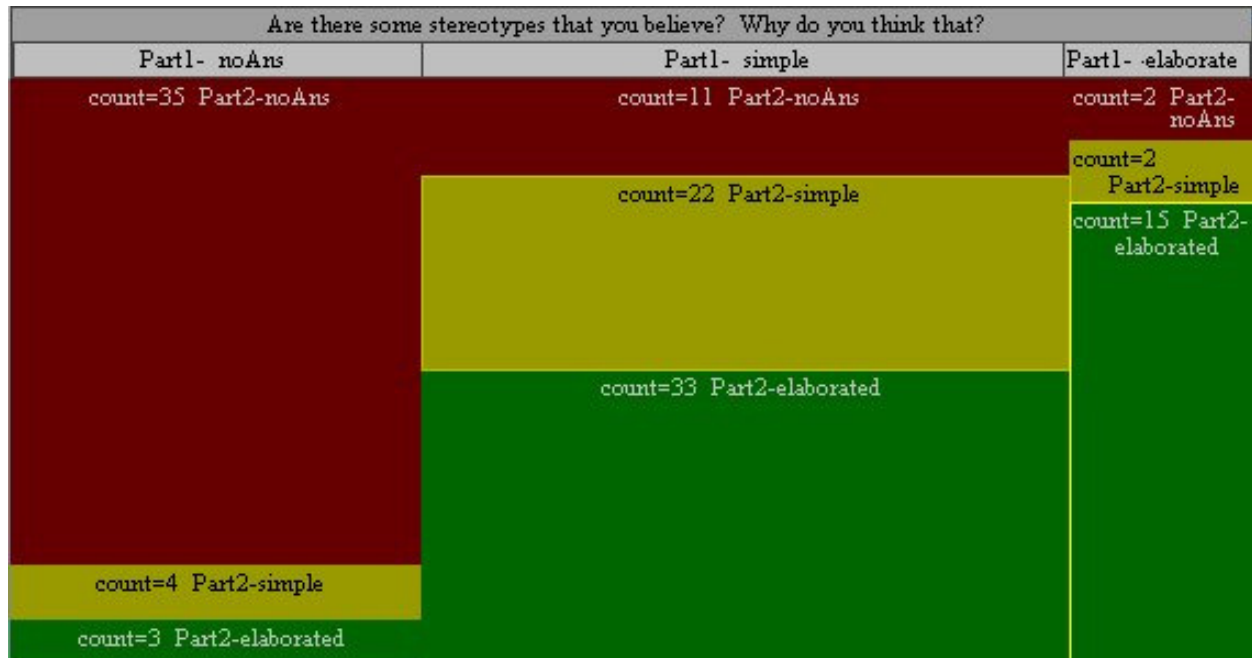


Figure 5. Treemap for “Are there some stereotypes that you believe? Why do you think that some stereotypes are true?”



For a 2-Part question, is a more general follow-up question better than a specific follow-up question?

We compared the responses to the following questions:

<b>General Follow-up</b>	DAISY > Are there some stereotypes that you believe? DAISY > Why do you think that?
<b>Specific Follow-up</b>	DAISY > Are there some stereotypes that you believe? DAISY > Why do you think that some stereotypes are true?

We found a statistically significant difference in visitors' responses to the second part of a 2-Part question between a general and specific prompt,  $\chi^2(2, N = 265) = 8.15, p = .017$ . See Table 10. Visitors were more likely to give an elaborated answer when a more general prompt was used.

**Table 10. Comparison of General versus Specific Follow-up for 2-Part Questions – Tally of types of answers according to type of question format**

Same Part 1 of 2-Part	DAISY > Are there some stereotypes that you believe?	No Elaborated Answer		Elaborated Answer	Total
		No Answer	Simple Answer		
<b>General Follow-up</b>	DAISY > Why do you think that?	64 (46%)	41 (30%)	33 (24%)	138
<b>Specific Follow-up</b>	DAISY > Why do you think that some stereotypes are true?	48 (38%)	28 (22%)	51 (40%)	127

This echoes findings from Davis' study on the effectiveness of generic versus directed prompts in encouraging reflection during problem-based science instruction<sup>6</sup>, we suspect for similar reasons: A more general follow-up prompt allows visitors to more broadly interpret Daisy's canned question, whereas a more specific prompt assumes visitors answered a certain way to the preceding prompt. If they had not, the conversation cannot proceed further. The following is an example of this difference:

With General Follow-up Prompt

```
DAISY> Are there some stereotypes that you believe?
ANWH,CEYRFWQVEQ> no i do not believe stereotypes.
/coded as simple answer/
DAISY> Why do you think that?
ANWH,CEYRFWQVEQ> you must think for yourself.
/coded as elaborated answer/
```

<sup>6</sup> Davis, E. A. (2003). Prompting middle school science students for productive reflection: Generic and directed prompts. *The Journal of the Learning Sciences*, 12(1), 91-142.

With Specific Follow-up Prompt

DAISY> Are there some stereotypes that you believe?

ELI> no

/coded as simple answer/

DAISY> Why do you think that some stereotypes are true?

ELI> i don't

/coded as simple answer/

Is a question that asks directly about the visitor's thinking better at fostering self-reflection than one that asks about machine thinking?

We compared two very similar questions that focus on the nature of thinking. The two, however, differs in the perspective taken, the machine asking the visitor to justify s/he's human and the machine asking the visitor to explain why the computer is (not) human:

<b>Human is the subject</b>	DAISY > How do I know I'm talking to a human and not just another machine?
<b>Daisy is then subject</b>	DAISY > Are you sure that I'm not a real person talking to you by e-mail? DAISY > What would it take to convince you?

We note that this is not a rigorous comparison because two different question formats were used. Nonetheless, it provides us is some indication of which may be better at eliciting self-reflective responses at Daisy.

We found that when the visitor is the subject, visitors were more likely to give an elaborated answer about themselves. Fisher's Exact Test,  $p < 0.0001$ . This suggests that we should **include questions that ask visitors pointedly about themselves in order to promote self-reflection.**

**Table 11. Comparison of General versus Specific Follow-up for 2-Part Questions – Tally of types of answers according to type of question format**

	Question	Elaborated Answer		Total Elaborated Answer
		Self-Reflection	Other	
<b>Human is the subject</b>	DAISY > How do I know I'm talking to a human and not just another machine?	32 (23%)	7 (5%)	39
<b>Daisy is then subject</b>	DAISY > Are you sure that I'm not a real person talking to you by e-mail? DAISY > What would it take to convince you?	8 (6%)	55 (38%)	63

## RECOMMENDATIONS

Based on the findings from this study and the previous study, we recommend that Daisy includes the following canned questions for the accompanying reasons. (See Table 12.)

**Table 12. Suggested canned questions for Daisy**

Previous Iteration	Proposed Final Version	Rationale
DAISY > Are you usually a logical person? DAISY > Give me an example.	DAISY > Are you usually a logical person? DAISY > Give an example of a recent logical or emotional decision you made.	There was less confusion with the second version. That is; less visitors were confused about what Daisy meant by 'an example'
DAISY > Are you sure that I'm not a real person talking to you by e-mail? DAISY > What would it take to convince you?	DAISY > How do I know I'm talking to a human and not just another machine?	This question was better at getting people to talk about themselves and not just Daisy.
DAISY > My mind works by recombining and regurgitating things that people tell me. But how is that different than what your mind does?	DAISY > My mind works by recombining and regurgitating things that people tell me. Doesn't your mind do the same thing? DAISY > In what ways are our minds different?	This worked slightly better as a 2-Part question.
DAISY > Are there some stereotypes that you believe? DAISY > Why do you think that some stereotypes are true?	DAISY > Are there some stereotypes that you believe? DAISY > Why do you think that?	A more general follow-up in this case allowed people who responded 'no' to the first part to, nonetheless, respond to the second part.
DAISY > Are you sure that I'm not a real person talking to you by e-mail?	DAISY > Are you sure that I'm not a real person talking to you by e-mail? DAISY > What would it take to convince you?	Adding a follow-up question that asks for an elaboration helped elicit more answers
DAISY > If I say something that makes sense, does that mean I was thinking? DAISY > If I wasn't thinking, where did the thought come from?	DAISY > If I say something that makes sense, does that mean I was thinking? DAISY > Where did my thoughts come from?	A more general follow-up allowed people who responded 'yes' to the first part to continue with the conversation.

**ACKNOWLEDGEMENTS**

The author would like to thank Alyssa Freedman for help in coding the computer logs.

This material is based upon work supported by the National Science Foundation under Grant number 0307927. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

