MORI Auto-Coder Guide

~ A Living Document ~

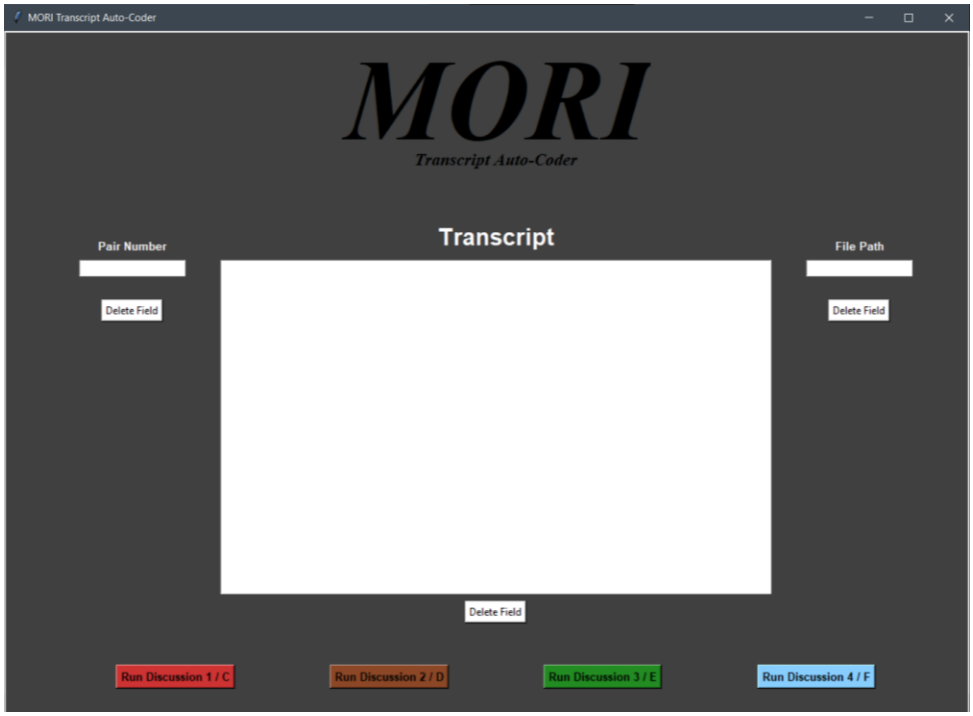
Contents

How To Use The Auto-Coder

The Auto-Coder is GUI based python file that auto-codes MORI transcripts, and can be run from any operating system that can run python.

.xlsx files are the only type of file supported, and (unless you want to tinker with the code) you must use the one provided, as it is prepared specifically for this coding.

When opened, it will look something like this:



Entry Boxes

Under the *Pair Number* heading is a box. This is where you will put in the number for the participant pair.

Under the *Transcript* heading is a box. This is where you will copy/ paste your correctly formatted transcript. Note that right clicking in the entry box does not work, but **Ctrl+C** and **Ctrl+V** does. You should be able to scroll through your transcript.

Under the *File Path* heading is a box. This is where you will put your file path to the designated excel sheet. If your Excel sheets are in a different location from the Auto-Coder .py file, then you will need to paste in the entire file path. For example:
C:\Users\your_name\Desktop\session_1.xlsx

However, it is much easier if the .py file is in the same location as the excel sheet (for example, on your Desktop, or in Documents). If they are, then you can simply type in the name of your Excel sheet into the box, without the file path:
code test.xlsx

Once finished with the 3 entry boxes, click one of the buttons on the bottom of the page. If it is Discussion 1/Discussion C that you are attempting to analyze, press the "Run Discussion 1/C" button. Then, check your excel sheet.

NOTE: You do **NOT** need to close the program down after each run. To delete the text in the box, select the *Delete Field* button. You can simply paste in a new transcript, file, and pair number, and then press a discussion button.

The Excel File

The Excel file is an important and necessary part of the auto-coder.

ALL blank excel files will look something like this:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	pair	discussion	type	detail	a_correct	b_correct	a_response	b_response	a_discussion_typ	b_discussion_typ	a_misled	b_misled	a_exposed	b_exposed	a_conf	b_conf	a_order	b_order
2			Same discussed	graffiti	ers	ers												
3			Same discussed	lockers	blue	blue												
4			Same discussed	warning	no mobile phones	no mobile phones												
5			Same discussed	stolen_tech	laptop	laptop												
6			Same discussed	appliance	microwave	microwave												
7			Diff discussed	utensil	sieve	frying pan												
8			Diff discussed	letter	t	k												
9			Diff discussed	jumper	blue	black												
10			Diff discussed	table_item	mug	bowl												
11			Diff discussed	TV_screen	car advert	abstract pattern												
12																		

Do not worry about the details – they will be changed by the auto-coder. At the bottom of the excel file, make sure that the sheets are intact:

Data

The ***pair*** column will be filled by user input in the GUI.

The ***discussion*** column will be filled by the GUI, dependent upon which Discussion button is pressed.

The ***type***, ***detail***, ***a_correct***, and ***b_correct*** will be automatically changed to the necessary details by the excel sheet (RefSheet) when the ***discussion*** column is filled.

All other columns will be filled by the auto-coder.

If all is completed correctly, then it will look something like this:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	pair	discussion	type	detail	a_correct	b_correct	a_response	b_response	a_discussion_typ	b_discussion_typ	a_misled	b_misled	a_exposed	b_exposed	a_conf	b_conf	a_order	b_order
2	5	C	Same discussed	elevator	blue	blue	blue	blue	3	3	2	2	7	5	3	2	2	1
3	5	C	Same discussed	time	11:45	11:45	11:45	11:45	3	3	2	2	0	5	2	3	1	2
4	5	C	Same discussed	accessory	glove	glove	glove	glove	3	3	2	2	0	5	3	3	1	2
5	5	C	Same discussed	sign	cafe	cafe	cafe	cafe	3	3	2	2	5	0	3	3	2	1
6	5	C	Same discussed	laptop_screen	mountains	mountains	mountains	mountains	3	3	2	2	5	5	3	3	2	1
7	5	C	Diff discussed	drawing	house	tree	tree	tree	2	1	1	2	1	0	3	3	2	1
8	5	C	Diff discussed	landmark	eiffel tower	leaning tower of pisa	leaning tower of pisa	leaning tower of pisa	2	1	1	2	1	3	3	1	2	1
9	5	C	Diff discussed	buttons	3	1	5	5	4	4	1	1	2	2	3	3	1	2
10	5	C	Diff discussed	cutlery	metal spoon	plastic fork	plastic fork	plastic fork	2	1	1	2	1	3	3	3	2	1
11	5	C	Diff discussed	money	cash	credit card	cash	credit card	5a	5a	2	2	1	1	3	3	1	2

Transcript Format

Begin with "question X."

Begin the next sentences with Experimenter, Person 1, or Person 2

Lowercase everything except for:

The options: A,B,C,D,E

The answers: F,K,M,G,T

The answers: ERS, EWM, SDL, CAW, SAR

***If your sentence begins with A, like "A glove was on the ground" then the 'A' should be lower case: "person 1: a glove was on the ground**

question 1.

experimenter: What was the colour of the sky?

person 2: i want to say it was blue or green.

person 1: i think it was blue.

person 2: Ok i - i feel like it was, sure yea let's go with blue.

experimenter: ok so both of you guys are going with blue there?

person 1: yes

person 2: (thumbs up)

P1 conf: h.

P2 conf: m.

Confidence Format – h/m/l

P1 F: D.

P2 F: D.

Final Answer Format

~

Question Ends with tilde (~)

Plain Example

question 1.

person 2: a glove was on the ground, so I think A.

person 1: i think it was a watch.

person 2: ok i - i feel like it was, sure yea let’s go with watch, I guess.

experimenter: ok so both of you guys are going with D, watch there?

person 1: yes.

person 2: (thumbs up).

p1 conf: h.

p2 conf: m.

p1 F: D.

p2 F: D.

~

‘But How Do It Know?’

The type of information the autocoder receives can be considered as coming from two sources: Transcriber perception, and conversational details.
Transcriber Perception is the confidence and final answer portion of the transcript. These parts must be completed by the transcriber and *can* differ via interrater error.
Conversational Details are the rest of the conversation. This includes *discussion type, misled status, misinformation, exposure type,* etc.

question 1

person 2: a glove was on the ground, so I think A

person 1: i think it was a watch

p1 conf: h

p2 conf: m

p1 F: D

p2 F: D

~

What Happens Behind The GUI?

Let’s walk through the process.

1. The entire transcript is run through, and every letter (except ones designated by the program, such as the options and capital letter answers) are lowercased. As well, all of the exclamation points and question marks are changed to a period.

2. The entire discussion is changed from a string to a list, where the 16 questions are sliced by the tilda (~) mark.

3. The program then grabs the first index, and if it contains ‘question 1’ it continues.

4. It then takes that first index, and slices it into another list via new paragraph (\n), to iterate through.

5. The program then removes all sentences that contain the word ‘experimenter’.

6. If person 1 is in the correct index it designates person 1/2 as the first speaker.

7. It then looks at exposure: if, for example, ‘person 1’ and ‘taj’ are in any index, then the program identifies that person 2 was exposed to the mention of ‘taj mahal’ and then appends the option to a new list entitled ‘MisinfoP2’.

8. The program then computes a new variable called ‘FinalMisinfoP1 / 2’ which is the exposures that a participant had, minus their witnessed detail.

9. It then computes ‘check’ which is used in a large if-else statement that computes the results of all exposures.

10. The program then finds the confidence and final answers via indexing the list for ‘p1 F’ and ‘p1 conf’

11. It then computes 3 misled types

12. And it computes 10 discussion types

13. Finally, at the end of the question, it computes the total misinformation exposures via a list, which is then extended throughout the rest of the questions in the discussion.

14. It does this for all of the questions in a discussion, and populates the excel sheet with the correct answers, as well as prints out more clarifying statements in the GUI.

Helpful Tips

1. While only specific things need to be capitalized/ not capitalized, you should try to lowercase everything that doesn't have to be, to reduce the chance of a type I error. As well, to reduce the chance of capitalization errors you should disable auto-capitalization in your word processing software.

2. Do **NOT** transcribe question 0 (the practice question).

3. Numbers can be written as words (e.g., three) or numerals (e.g., 3) – they are not critical details and are not analyzed.

4. You may end lines for each person with periods (and other punctuation), and can have periods in the middle of lines for each person (i.e., you can transcribe speech as it would be written).

5. Leave blank any confidence ratings/final answers that aren't given (shouldn't actually exist though)

Troubleshooting

In the likely event something went wrong, it's most likely the tedious format.

I think some of the Misled or Exposures are wrong.

Are the grafitti, the letters, time, or the answers in capitals/ numerical format?

a_response / b_response isn't showing up / correct.

Is it formatted correctly, like this? P1 F: A.

P2 F: A.

Is there a tilda (~) separating each question?

a_conf / b_conf isn't showing up / correct.

Is it formatted correctly, like this? P1 conf: m.

P2 conf: h.

Is there a tilda (~) separating each question?

Speaker order is wrong

Is it formatted correctly? Does it begin with question X? Is there a tilda (~) separating each question?

----------THE CODING SCHEME ----------

The following section details the coding scheme that the auto-coder uses

The following variables are coded by the Auto-Coder program using formatted transcripts:

Discussion type:

Categorizes the pair's agreed upon final answer. A response is considered a final answer when:

1. Both participants explicitly declare the same final answer (e.g., "final answer is blue")
2. One participant explicitly declares a final answer on behalf of the pair (e.g., "our final answer is blue") and the partner appears to agree
3. When prompted for a final answer, both participants declare the same answer (e.g., "blue")
4. When prompted for a final answer, one participant declares an answer on behalf of the pair (e.g., "blue") and the partner appears to agree
5. Both participants explicitly declare different final answers (e.g., "final answer red" and "final answer purple")
6. When prompted for a final answer, both participants declare a different answer (e.g., "blue" and "purple")

Note: all other scenarios will be coded as "no response" under 5(f) below. Points 5 and 6 will be coded as pair provided separate answers under point 5 below.

Each of the 10 discussion items will be categorised into one of the following categories:

Final answer

1. Pair agreed on **participant's** witnessed detail
2. Pair agreed on **partner's** witnessed detail
3. Pair agreed on **joint** witnessed detail
4. Pair agreed on other **inaccurate** detail (e.g. detail on the 5AFC that neither participant saw or detail that was not mentioned on the 5AFC)
5. Pair provided separate answers
    (a) participant provided participant-witnessed detail
    (b) participant provided partner-witnessed detail
    (c) participant provided other inaccurate detail mentioned by partner (detail mentioned on the 5AFC that neither participant saw, or detail that was not mentioned on the 5AFC)
    (d) participant provided other inaccurate detail NOT mentioned by partner (detail mentioned on the 5AFC that neither participant saw, or detail that was not mentioned on the 5AFC)
    (e) participant provided non-relevant detail
    (f) participant provided no response

Exposure to misinformation

Step 5: For each discussion item and each participant, the Auto-Coder program will code whether the participant was exposed to misinformation **at any point during** the discussion.

Exposure to misinformation

1. Non-discussed (not exposed to any information, auto-coded)
2. Exposed to the partner's witnessed detail
3. Exposed to an incorrect detail that neither participant witnessed
4. Exposed to the participant's witnessed detail
5. Both 1 & 2
6. Both 1 & 3
7. Both 2 & 3
8. All of 1, 2, & 3

If a person mentions a detail, but does so in the negative, e.g., Person A says "I *didn't* see a glove", and "glove" is incorrect for both Person A and B, it still counts as exposure to misinformation (i.e., a code of 2).

Misled status

For each discussion item, the Auto-Coder will determine whether or not a participant was misled by their partner.

Participants will be coded as being **misled by their partner** using the following algorithm:

1. Check participant's final answer—is it incorrect? If yes, proceed.
2. Check discussion—was incorrect final answer mentioned explicitly by partner? If yes, misled.
3. Otherwise, not misled.

Note: If a participant is exposed to more than one detail, both details will be used to determine whether or not a participant was misled by their partner (e.g., if a participant is exposed to both a partner-witnessed detail and an inaccurate-detail, the partner will be considered to have been misled if they have agreed on the partner-witnessed detail OR the inaccurate detail)

Note (in cases where a pair agrees on an inaccurate final answer): If, before a pair declares a final answer, a participant does not explicitly mention an inaccurate detail but agrees with an inaccurate detail mentioned by their partner, this participant will be coded as "misled" and their partner will be coded as "not misled".

e.g., A reports "5 buttons" (incorrect for A, incorrect for B), B response with "I agree", A is not misled (Because B did not explicitly mention "5 buttons"), B is misled

e.g., A reports "5 buttons" (incorrect for A, incorrect for B), B response with "Yes, 5 buttons", both are misled

e.g., A reports "5 buttons" (correct for B, incorrect for A), B reports "3 buttons" (correct for A, incorrect for B), A agrees with "3 buttons", B agrees with "3 buttons". A is not misled, B is misled (Because A explicitly mentioned "3 buttons")

All other scenarios will be coded as being **not misled by their partner**.

Speaker order

For each discussion item for each participant, the Auto-Coder will determine whether the participant spoke first or not:

Order of speech
1. Participant spoke first
2. Partner spoke first

Step 7: For each discussion item for each participant, the transcribers will rate how confident the participant appears in their memory of each item.

Confidence
How confident does the participant appear to be in their memory of the detail:

1. Not at all confident
2. Moderately confident
3. Very confident

We considered the use of more objective criteria (e.g., certain verbal/non-verbal cues), but based on previous experiences trying to do this we stuck with subjective ratings.