Codebook

Note:

- Subtopics contained in the task description are: (1) the notion of gut-brain connection, (2) underlying mechanisms, (3) gut microbiome—physical health, (4) gut microbiome—mental health, (5) factors influencing gut health, and (6) ways to improve gut health.
- By default, one row will be associated with one code, with the exception of 'building' and 'planning' (e.g., when someone creates a list of subtopics at the beginning of the task upon reading the task description). For the instances where two codes can cooccur, C1[1] will try separating one macro activity into two micro activities (two rows) to minimize ambiguity.

Search activities: Typically observed when a participant interacts with a search engine; search activities here refer to the actions a person takes when deciding what to search for, formulating a query and evaluating the search results.

1 Query formulation

1.1 **Structure-driven** (also known as top-down): Query formulation is guided by the structure of the topic (i.e., titles of subtopics). Typically, a participant would copy and paste a statement from the task description or formulate their own query using the key words contained in the task description. Structure-driven search can also happen when a participant decides to gather additional information on a (sub)topic they had already searched in a previous iteration that is part of the task description, not something they encountered during their search.

1.2 **Data-driven** (also known as bottom-up): Query formulation is guided by the data encountered during a search. For instance, when someone reads about the gut-brain connection and comes across the concept of blood-brain barrier (BBB), conducting an additional search to understand the meaning of BBB is considered a data-driven search. In a different scenario, someone reads an article that mentions a study involving patients with schizophrenia and conduct a search to follow up on the relationship between gut microbiome and schizophrenia, this is also considered a data-driven search. It is likely that a search query that is not directly derived from the task description (a given set of subtopics) will be considered a data-driven query; such a query tends to be more specific than a structure-driven query (e.g., 'does the gut microbiota influence schizophrenia' vs. 'gut microbiota and mental health').

2 Search strategy

2.1 **Being selective with source**: A participant is mindful about the sources during search results evaluation, considering factors such as credibility, reliability, or intended audience. For instance, we apply this code when a participant notices and comments on specific aspects of the source, such as the name, the source's credibility, the author's expertise, or whether the article is meant for a general audience or scholars. This also includes an instance where the participant

[1] C1: coder 1

(re)formulates a query to target a specific source (e.g., Wiki, Amazon, Reddit). This code only applies for document-level judgments, not passage-level.

2.2 **Searching for structure**: A participant seeks information that will help them gain a high-level overview of the topic. High-level understanding here refers to the kinds of things that are commonly discussed regarding the topic (gut-brain connection). For instance, if someone looks up the table of contents of a book with the intent to see how a discussion around the topic can be structured, we code this as 'searching for structure'.

Sensemaking activities: Typically observed when a participant interacts with their notes, which we consider as a knowledge structure. However, certain activities, such as gap identification and assessing semantic fit, may occur at any point in the process; when a person scans a SERP, reads the content of an article, reviews their notes, or tries to understand how new information can be integrated into the existing knowledge.

1 **Accretion**: A participant adds relevant information into their notes. They might do so by copying and pasting text, writing down key terms or phrases, rephrasing the text, or summarizing the text into a shorter and more accessible form.

2 Structuring

2.1 **Building**: A participant creates a structure in their notes. This can manifest as them adding (sub)headings, using a bulleted list or creating paragraph breaks.

2.2 **Tuning**: A participant modifies the existing structure of their notes. This can manifest as them changing the name of the (sub)headings or using formatting (bold, color, or highlight) to assign weight to a specific concept. The idea behind this code is that as a person engages with and absorbs information during their search, their understanding of the topic becomes more refined, which leads them to update their notes to align with their evolving representation of the topic. This also involves a person arranging the location of information within a subtopic (to make it flow better). This happens when one has more than a couple of bullet points underneath one heading and they are trying to organize those bullet points in a sensible way. Changing the name of the (sub)headings should not be confused with restructuring unless the name change is due to the structural change of the person's understanding of the topic. For example, if someone changes a heading from 'connection between gut and brain' to 'bi-directional communication system between gut and brain', that would be tuning. If someone changes a heading from 'connection between gut and brain' and changes it so that now they have two subheadings, 'gut → brain' and 'brain → gut', that would be restructuring.

2.3 **Instantiating**: A participant adds more detailed information about key concepts or (sub)topics. This can manifest as them creating sub-bullets within a list, adding specific examples related to the abstract concepts, or elaborating on previously written statements by incorporating new, detailed information into the existing text.

2.4 **Restructuring**: A participant revises the existing structure in their notes by separating, merging, or deleting the (sub)headings or paragraphs of the text.

3 **Gap identification**: A participant identifies a gap in their knowledge. In our data, this can manifest in

the following ways: (1) admitting a lack of familiarity with an encountered concept (e.g., "I have no idea what endocrine system is"); (2) posing a question verbally (e.g., "what is endocrine system?"); or (3) indicating the gap in their knowledge in their notes (e.g., write down 'endocrine system??', 'endocrine system (search)' or writing it down at the end of the notes as a place holder). The last scenario is different from building. We don't consider it as building until the newly encountered information is pursued and gets incorporated into the existing knowledge structure. Until then, the note-to-self functions more as a to-do item or a reminder to fill in the gap.

4 **Semantic fit**: A participant evaluates whether the new information encountered fits into the knowledge structure or meets the information need (goals) in mind. This can manifest as participants questioning how the encountered information relates to the larger topic or rearranging statements within their notes to place them under the most suitable subtopic. The bottom line is, they are trying to evaluate whether the data fits into their existing mental representation of the topic and, if so, where it fits best. If someone doesn't know where the information fits because they don't understand the information to begin with, we code it as 'gap identification' instead of 'semantic fit' even though the fit was also in question. Another evidence of this is when a participant can switch between different bins seamlessly while reading through the article or when they start recording an information in one bin than move it to another bin before finishing their thoughts. These all show that the person is thinking about which bin the current information fits the best and they can do so because they have subgoals active in mind. In that sense, 'semantic fit' and 'active maintenance' have conceptual overlap, but 'semantic fit' focuses on keeping the (sub)goals active in mind whereas 'active maintenance' focuses on keeping the previously read information active in mind.

Cognitive activities

1 **Active maintenance**: A participant keeps certain information active in their mind and use it to either (1) keep themselves on track not going into the rabbit hole (e.g., saying aloud the current goal to remind themselves where they should be), (2) avoid redundancy (e.g., avoiding the redundant information or the search result), or (3) realize how new information is related to previously encountered information (what they've read or written out up to the present moment). The third scenario includes the participant noticing the similarities (or differences) between the content.

2 **Planning**: A participant engages in planning activities during their search session. This may manifest in two ways: (1) a participate perform a task analysis at the beginning of the study session, meaning they spend a couple of minutes re-reading the task description to decide where to begin, what to search for in what order, etc. or (2) a participant verbalizes their plan for the next iteration of search (e.g., "I'm going to look up X after this", "I'm looking up X")

3 **Monitoring**: A participant monitors their working progress. This may manifest in two ways: (1) a participant refers to the task description or their own notes to see where they are at in terms of covering the different, suggested subtopics (essentially, annotate anytime they check in with the task description or scan notes to see where they are at, almost like checking your current location on the map) or (2) a participant checks in with themselves or with the study moderator to assess how they want to use the remaining time.

4 **Reflecting**: A participant stops searching and enters a reflective phase where they go over gathered

information again and consolidate their learning. For instance, a participant would pause their search to go over and synthesize their notes periodically throughout the session or they would stop their search, shift their attention to their notes, and dedicate time to examine the gathered information, ensuring a solid understanding of the concepts learned by the end of the task. This does not necessarily have to happen at the end of the task.

Sources of potential confusion:

- Gap identification vs. Semantic fit: Although these two activities may have an overlap, they are not the same. Gap identification can be understood as a precedent of semantic fit. For instance, if someone does not know what 'endocrine system' is, the need to know what it is first may precede the need to understand if and how it has anything to do with the larger topic they are asked to learn about. In another scenario, a person encounters 'endocrine system' and although they are not familiar with the concept, they kept on reading and are able to quickly gain a basic understanding of what it is and only questioned its relations to the larger topic, in which case we observed not gap identification but semantic fit.
- Monitoring vs. Reflecting: Monitoring involves checking the current status in relation to the overall goal, task progress, and resource management (e.g., what's left to do, time left). Reflecting, on the other hand, is a different process, where a person reviews the information gathered and ensures comprehension of the concepts. While monitoring guides decisions for the next steps (e.g., what to search for next, when to move on), reflecting involves activities such as organizing notes, rephrasing, and synthesizing the information learned, leading to a more solidified understanding of the topic. Reflecting typically takes more time than monitoring too.
- Building vs. Planning: Building necessarily entails note-taking activities, while planning can be expressed verbally. With that being said, there can be instances where both codes are applicable. For instance, when someone creates a bulleted list consisting of subtopics upon reading the task description, this can be coded as both 'building' and 'planning' assuming that they created the list with the intent to touch on the subtopics included in the list throughout the session. These cases will be represented by two rows associated with one activity and we add the 'building' code first before the 'planning'.
- Accretion vs. Instantiating: By default, when a participant adds a new piece of information to their notes, we annotate 'accretion'. Instantiating happens when the participant builds on what has been already recorded (in other words, what they have learned in previous search iterations) by either elaborating on a point, adding examples relevant to the point, or adding just more information that supports the existing point in some way.
- Copying and pasting a link to a source → We count it as neither planning nor accretion.