

MTurk ID Collection

What is your MTurk ID?

Instructions 2

Welcome to our survey!

In this survey, you will be completing a pre-test that allows you to earn a qualification for a highly-paid future task. In the future task, you will be reading about an activity that an individual or group performs. The task will then ask you to rate the activity by answering 23 questions about it.

This survey has 2 parts:

1. You will answer **6 training questions** (selected from the total pool of 23), in which we will provide you with the correct answer after you respond. You should do your best to read and answer each question. Especially if you get a question wrong, please read the feedback and make sure that you understand any possible misunderstandings.
2. You will answer **16 graded testing questions**. We will NOT share the correct answers for these questions. Your score on these questions will determine whether you become eligible for the future task.

You should budget approximately **20 minutes** to complete both parts.

We want to make sure our instructions make sense. Please answer the following 4 questions before proceeding to the pre-test.

(Question 1 of 4) How many training questions are there?

(Question 2 of 4) How many graded testing questions are there?

(Question 3 of 4) How many total questions are there in the future task?

(Question 4 of 4) What is the purpose of the future task?

To get familiar with a future task and understand the most challenging questions

To work on an activity as an individual or in a group

To ask people to make decisions as a group

To rate a task performed by individuals or groups by answering a number of questions

Training

You will now see **6 training questions**. First, you will read about a description of a task or activity. Then, we will show you 6 questions about it, which are selected from the total set of 23.

On each page, you will see the following:

- A reminder of the task or activity
- The question
- An elaboration of the question, with additional examples to help you answer
- The answer choices.

Please read the task description and then answer the question. You will receive feedback after answering these questions.

Press the 'next' button to proceed.

The name of the task you will be focusing on is called the **Word completion given starting letter** task. Please read about the task below. You will also see a reminder of this description on the page of every question.

When you are done reading, please advance to the next page to answer the first training question.

Word completion given starting letter

1. Set-Up

Participants are given a starting letter and ending letter for words. For example, the start letter could be "S" and the end letter could be "N." The starting and ending letters are selected such that there is always a nonzero set of English words that can be created.

2. Objective / Goal

The goal is to generate a list of valid English words using the starting and ending letter provided. Using the previous example, participants could be asked to generate words like "SPIN," which starts with "S" and ends with "N." Participants must generate the maximum number of valid words they can within a fixed amount of time. Nonvalid words will be rejected by the system.

Here is a reminder of the task:

Word completion given starting letter

1. Set-Up

Participants are given a starting letter and ending letter for words. For example, the start letter could be "S" and the end letter could be "N." The starting and ending letters are selected such that there is always a nonzero set of English words that can be created.

2. Objective / Goal

The goal is to generate a list of valid English words using the starting and ending letter provided. Using the previous example, participants could be asked to generate words like "SPIN," which starts with "S" and ends with "N." Participants

must generate the maximum number of valid words they can within a fixed amount of time. Nonvalid words will be rejected by the system.

QUESTION 1 (of 6). Is the goal (or one of the goals) of this task to try to achieve a precise outcome?

Elaboration: Here, there is a specific standard to meet, but precision matters.

Look for tasks with a specific, most preferred outcome. If the task is to estimate the temperature of a room, the goal is to agree with the value indicated by the thermometer. Another example is moving exactly 10 objects --- no more, no less --- or to exactly reproduce something (such as a task where people have to exactly copy a work of art).

Also look out for concepts like "exactly," "precisely," "optimal," or "best." For example, if you are asked to generate exactly 10 ideas, generating both 9 ideas and 11 ideas would be considered a failure. Another clue might be terms like "constraints," "requirements," or "rules" that someone has to follow, as long as there is a specific best answer within the rule set. For example, if you want to buy the most number of items using a budget of \$100, the "constraint" of \$100 also serves as a precise goal. That is, the closer you get to \$100 (the more precise you are), the better you do at the task.

No, one of the goals is NOT to achieve a precise outcome.

Yes, one of the goals is to achieve a precise outcome.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

The answer to this question is **no**. The instructions say to generate "the maximum number of valid words" --- and there is no exact amount. Additionally,

any valid words count; you are not being able to generate a specific set of words. Thus, we should answer no.

If you got this question wrong, you might wonder whether the starting and ending letters count as "rules" or "constraints." But remember, there needs to be a "best" answer in the set of rules that the designers are looking for, and that is not the case here.

HANDY TIP: When answering this question, focus on the outcome or goal. What are participants being asked to do? Are they trying to get as close as possible to the "right" or "best" answer, or just trying to do their best? In this case, the goal is oriented towards producing as many words as possible, and not towards precision. (In fact, even if you are imprecise, it's fine; according to the description, the system will automatically reject any invalid words.)

As another example, consider the difference between this task ("generate as many words as you can that start with S and end with N") and a task that says, "generate a 3-letter word that starts with S, ends with N, and is also the name of a star." Adding this third condition makes the task a precision task, because the person writing the task is now looking for a specific word: SUN. This task, on the other hand, does not have a precise outcome, since the people writing the task are not looking for any word in particular.

Here is a reminder of the task:

Word completion given starting letter

1. Set-Up

Participants are given a starting letter and ending letter for words. For example, the start letter could be "S" and the end letter could be "N." The starting and ending letters are selected such that there is always a nonzero set of English words that can be created.

2. Objective / Goal

The goal is to generate a list of valid English words using the starting and ending letter provided. Using the previous example, participants could be asked to generate words like "SPIN," which starts with "S" and ends with "N." Participants

must generate the maximum number of valid words they can within a fixed amount of time. Nonvalid words will be rejected by the system.

QUESTION 2 (of 6). Is there only one method for achieving the task, as opposed to many alternatives for task completion?

Elaboration: Is there only one right way of solving the problem?

Answer 'no' if there is more than one possible course of action or process to attain the group's goal.

Answer 'yes' if there is only ONE process or action that will lead to the correct answer or achieve the goal.

As an example, answer 'yes' if the instructions specify exact steps, e.g., "first you write all the 3-letter words, then the 4-letter words, then the 5-letter words," or if there's really only one way to do the task (e.g., perform long division). Actions might also be limited by the environment: participants may be working on an electronic system that restricts their communication or the steps that they are allowed to take.

However, answer 'no' if the task is open-ended in terms of course of action: e.g., if participants are simply asked to "come up with as solution" or "give as many ideas as possible" without specifying how they should achieve the goal.

No, there are MANY ways to achieve this task.

Yes, there is only ONE way to achieve this task.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

The answer to this question is **no**. There are many ways of completing this task, and the instructions do not specify exactly how to do it, so long as you generate as many words as possible.

Of course, there are some methods that will be harder to execute than others --- for example, if your method involves generating a large number of invalid words, the system will reject them --- but there are still tons of options! You could start with generating shorter words and then move on to longer words; you can just try to think of as many words as you can, regardless of length; you can even randomly choose letters of the alphabet and hope that they form valid words; and so on. There are many methods of doing this task!

Here is a reminder of the task:

Word completion given starting letter

1. Set-Up

Participants are given a starting letter and ending letter for words. For example, the start letter could be "S" and the end letter could be "N." The starting and ending letters are selected such that there is always a nonzero set of English words that can be created.

2. Objective / Goal

The goal is to generate a list of valid English words using the starting and ending letter provided. Using the previous example, participants could be asked to generate words like "SPIN," which starts with "S" and ends with "N." Participants must generate the maximum number of valid words they can within a fixed amount of time. Nonvalid words will be rejected by the system.

QUESTION 3 (of 6). Can this task be written as a "formal model" that an algorithm could solve?

Elaboration: Some tasks can be written as a formal model or math problem, expressed with rules and syntax that the problem-solvers share.

One way to think about this question is to ask yourself, could a robot or algorithm do this task?

Here are examples of a 'yes' answer:

1. A specific set of rules and outcomes: Robots are really good at following rules. Answer 'yes' if you can input the rules and the desired outcome, and have a robot follow predetermined steps to get the result. For example, you can tell the robot, "buy as many items as you can with a budget of \$100." (In fact, online shopping websites have exactly this tool!)
2. You can write it like a math problem: Robots are really good at doing math. Something like, "find the best teams under the constraints," combined with a list of how much "utility" people get from being put together, can easily be put into a computer and solved. Similarly, if you're trying to find the shortest path from Point A to Point B, you can think of this in terms of geometry.
3. You can break the problem down into small units of meaning: Examples of these units include the alphabet (26 letters), colors (which can be broken down into units of Red, Blue, and Green), or coordinates on a graph (X,Y). An algorithm could use these units of meaning to solve the task. For example, when generating words, the letters are the units. You can imagine an algorithm in which you use "brute force" to find every possible combination of letters that makes a valid word. Another example is Wordle --- many people have built bots that consistently solve the puzzle!

Here are examples of a 'no' answer:

1. Creative ideas that cannot be generated just by following basic rules. Whereas you can generate words just by using the rule of trying out all the different letters in the alphabet, you can't generate stories in the same way. There are no "rules" for what makes a good story. (We ask you not to think about advanced models like GPT-3 for the purposes of this question.)
2. Subjective tasks that require a judgment call. Deciding whether or not we should ban all guns can't really be done by an algorithm: people have to debate these ideas.

3. Tasks relying on social dynamics. Anything focusing on relationships between participants should be labeled 'no.'

No, this is NOT a formal problem that an algorithm could solve.

Yes, this is a formal problem that an algorithm could solve.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

The answer to this question is **yes**. We can represent all the words starting with "S" and ending with "N" like this:

S _____ N

In which we can fill the space with a combination of any of the 26 letters of the English alphabet. These 26 letters are the "units of meaning," and we can think of an algorithm that simply goes through and tries each letter, submitting the ones that are valid English words (SIN, SUN, SON ...). Then it can try combinations of two letters (SPIN, STUN, SPAN ...), combinations of three letters, and so on. You can easily imagine a computer program that can "brute force" this task and perform quite well!

Since it's possible to write this problem formally, in a way that an algorithm could solve, the answer to this question is yes.

Here is a reminder of the task:

Word completion given starting letter

1. Set-Up

Participants are given a starting letter and ending letter for words. For example, the start letter could be "S" and the end letter could be "N." The starting and

ending letters are selected such that there is always a nonzero set of English words that can be created.

2. Objective / Goal

The goal is to generate a list of valid English words using the starting and ending letter provided. Using the previous example, participants could be asked to generate words like "SPIN," which starts with "S" and ends with "N." Participants must generate the maximum number of valid words they can within a fixed amount of time. Nonvalid words will be rejected by the system.

QUESTION 4 (of 6). Is this a "generation" or "brainstorming" task? In other words, is one of the main purpose(s) of this task to produce a number of ideas or examples, without any particular action associated with them?

Elaboration: These are tasks based around generating ideas, examples, or concepts.

The examples being generated can be either abstract (generating words and colors) or concrete (generating ideas for how to spend \$10,000 or ways to use a paperclip). If the ideas are concrete, they should NOT suggest a specific plan or course of action.

Note that you should only answer "yes" to this question if one of the main outcomes of this task is to generate such ideas. Many tasks require people to discuss ideas (for example, a jury task), or come up with creative ways to solve a problem (for example, a math problem), but the primary goal of such tasks is not just to generate the ideas.

No, this is NOT a "generation" or "brainstorming" task.

Yes, this is a "generation" or "brainstorming" task.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

The answer to this question is **yes**. The goal of this task is EXACTLY to generate ideas, examples, or concepts --- the participants are generating examples of words that start and end with specific letters. Therefore, this task is a "generation" or "brainstorming" task.

Notice that the elaboration says that the "examples being generated can be either abstract ... or concrete." Words would qualify as being an example of generating abstract ideas. In the future, you may encounter tasks that ask the participants to generate something more concrete, such as ideas for an advertising campaign or different creative stories. All of these tasks would be a "yes" in this category.

Here is a reminder of the task:

Word completion given starting letter

1. Set-Up

Participants are given a starting letter and ending letter for words. For example, the start letter could be "S" and the end letter could be "N." The starting and ending letters are selected such that there is always a nonzero set of English words that can be created.

2. Objective / Goal

The goal is to generate a list of valid English words using the starting and ending letter provided. Using the previous example, participants could be asked to generate words like "SPIN," which starts with "S" and ends with "N." Participants must generate the maximum number of valid words they can within a fixed amount of time. Nonvalid words will be rejected by the system.

QUESTION 5 (of 6). Can the outcome of this task be described in win/lose terms?

Elaboration: This question captures whether a task is a "competition" or "battle," in which your outcome is relative to someone or something else (e.g., another player, team, or an AI agent).

A battle is "where the focus is on conquest of an opponent and winner-take-all distribution of payoffs," and a competition is "where there is a lot of emphasis on standards of performance excellence over and above the reckoning of winners and losers." Competitions between sports teams fit this category (e.g., basketball, soccer, ice hockey), but so too do military and street gang battles (e.g., the spoils of war go to the victor no matter how honorably the vanquished may have fought).

Answer 'yes' ONLY if you "win" or "lose" relative to somebody else who is playing the game. If someone performs poorly in the task, but they are not being compared to another player, you should answer 'no.' For example, if you are being asked to come up with as many ideas as possible, you may think that you "lose" if you generate zero ideas or that you "win" because you generated many ideas. However, you should still answer 'no,' because this question is asking about "winning" and "losing" in relative terms. A task in which you need to come up with more ideas than another player, on the other hand, would be a 'yes,' since you would win or lose relative to that other person.

No, the task is NOT a "competition" or "battle."

Yes, the task is a "competition" or "battle."

Not applicable or not answerable based on the task description (Please Elaborate Below.)

The answer to this question is **no**. Here, participants are NOT competing against anyone else --- they are simply trying to come up with as many words as possible.

The key definition in the elaboration is that "this question captures whether the task is a 'competition' or 'battle,'" in which you have to "win or lose **relative to someone else who is playing the game**." You might think, for example, that it is possible to "lose" this game if the participant does not come up with any ideas. However, as the elaboration clarifies, "this question is asking about 'winning' and

'losing' in relative terms." Since the description does not mention that there is an opponent, this task does not count as a competition or battle.

HANDY TIP: As you answer these questions, **pay close attention to what the goal of the task is, and how a participant is graded on the task.** This will often determine the correct answer. You can imagine a different version of this task, in which participants have to "race" against another player to generate more words within a specified period of time. In that case, you would be graded *relative to another player*, which means you would want to answer "yes" to a question like this one.

This tip will be useful for many questions. A different question that you will see in the graded quiz later is about whether a task is "all or nothing." You can imagine a version of this task that is all or nothing: for example, if participants get full credit if they generate 10 words, and no credit if they generate anything less, then it is all-or-nothing. You can also imagine a version of this task in which participants simply get points for every word they are able to generate, which would NOT be all-or-nothing. Paying attention to the details about how participants are graded will help you find the right answer.

Here is a reminder of the task:

Word completion given starting letter

1. Set-Up

Participants are given a starting letter and ending letter for words. For example, the start letter could be "S" and the end letter could be "N." The starting and ending letters are selected such that there is always a nonzero set of English words that can be created.

2. Objective / Goal

The goal is to generate a list of valid English words using the starting and ending letter provided. Using the previous example, participants could be asked to generate words like "SPIN," which starts with "S" and ends with "N." Participants must generate the maximum number of valid words they can within a fixed amount of time. Nonvalid words will be rejected by the system.

QUESTION 6 (of 6). Is it efficient and useful for members of the group to work on discrete parts (or subtasks) of this activity?

Elaboration: Ask yourself: is it efficient or useful for a team of people to "divide and conquer" this task, or is this really something where one person should be doing most of the work?

A key heuristic is whether the different sub-parts of the task are interdependent or not. If the sub-parts are not dependent on each other, it often makes sense for different people to work on each part separately (so answer 'yes'). However, if the sub-parts are interdependent, one person can't start their part without waiting for another person to finish, so dividing and conquering doesn't make sense (and you should answer 'no').

For example, if you are solving 10 simple arithmetic problems, each of ten people may work on one of them. If there is only one problem to be solved, it's not efficient for one person do the thinking and another do the writing. Similarly, for a task where you want to generate as many ideas as possible, everyone can separately come up with ideas and combine in the end. For a task where everyone needs to come up with a shared plan, the parts of the plan depend on each other, so it is difficult to divide and conquer.

No, you should NOT divide-and-conquer this task.

Yes, you should divide-and-conquer this task.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

The answer to this question is **yes**. Since the goal is to generate as many words as possible, it is a great idea to ask many people on a team to individually generate as many ideas as they can and then combine it in the end.

If you struggled with this question, think about whether you'd rather be doing this task alone or doing it with many people who can work separately before combining. In this case, a "brainstorming" task like this one will benefit a great deal from people putting their heads together!

Testing

You have finished the training portion of this survey!

You will now see **16 graded testing questions**. In this section, you will read about a description of two different tasks. Then, we will show you 8 questions about about each one; these questions are also selected from the total of 23.

Just as before, you will see the following on each page:

- A description or reminder of the task or activity
- The question
- An elaboration of the question, with additional examples to help you answer
- The answer choices.

Please read the task description and then answer the question. Remember that you will NOT receive feedback, but your qualification for the future task depends on your answers to these questions.

Press the 'next' button to proceed.

The first 8 questions will be about the **Shopping Plan Task**. Please read about this task below. As before, we will show you a reminder of the task on every page.

The first question for this task appears on the next page.

Shopping plan

1. Set-Up

Each participant has a grocery list and a map that shows distances and times between each grocery store, as well as a list of potential items to purchase and how many points they are worth.

2. Objective / Goal

Participants have to plan a shopping trip as though they were all residents of the same house sharing the same car. The goal is to purchase as many high-quality items as possible in a fixed amount of time, and to get the highest number of points by considering tradeoffs between price, quality, and driving time. Participants gain points for every item that they are able to plan to purchase. Getting anything less than the maximum number of points will lead to partial credit.

Here is a reminder of the task:

Shopping plan

1. Set-Up

Each participant has a grocery list and a map that shows distances and times between each grocery store, as well as a list of potential items to purchase and how many points they are worth.

2. Objective / Goal

Participants have to plan a shopping trip as though they were all residents of the same house sharing the same car. The goal is to purchase as many high-quality items as possible in a fixed amount of time, and to get the highest number of

points by considering tradeoffs between price, quality, and driving time. Participants gain points for every item that they are able to plan to purchase. Getting anything less than the maximum number of points will lead to partial credit.

QUESTION 1 (of 16). Is this a "planning" task? In other words, is one of the main purpose(s) of this task to produce a sequence of concrete steps or actions that an individual can follow to achieve some goal?

Elaboration: These are tasks "having emphasis on action orientation."

Answer 'yes' if this task involves writing down a series of concrete steps that someone will follow in order to achieve a goal. Examples include organizing an event, or figuring out a plan for buying things at the store. This plan may later be executed by the person making it, or could be given to someone else.

Answer 'no' if this task just involves doing an activity without writing the steps down. (For example, if you write a recipe and then cook, you are planning; if you cook without a recipe, that is NOT planning.) Also answer 'no' if there is no intention to actually use the plan in the "real world," or if it's just a hypothetical math problem. Sometimes math problems involve a story or context (e.g., "calculate how much time it takes for Jimmy to go to the store," or "find the shortest path that a traveling salesman can take". However, these are math problems rather than plans.

No, this is NOT a "planning" task.

Yes, this is a "planning" task.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

Here is a reminder of the task:

Shopping plan

1. Set-Up

Each participant has a grocery list and a map that shows distances and times between each grocery store, as well as a list of potential items to purchase and how many points they are worth.

2. Objective / Goal

Participants have to plan a shopping trip as though they were all residents of the same house sharing the same car. The goal is to purchase as many high-quality items as possible in a fixed amount of time, and to get the highest number of points by considering tradeoffs between price, quality, and driving time. Participants gain points for every item that they are able to plan to purchase. Getting anything less than the maximum number of points will lead to partial credit.

QUESTION 2 (of 16). Is a participant able to come up with a provably correct solution, assuming sufficient ability, time, motivation, and resources?

Elaboration: If someone is able to solve the problem (e.g., find the best solution), will that person "have sufficient ability, motivation, and time to demonstrate the correct solution to the incorrect members?" In other words: Assuming infinite time and resources, can someone both (1) solve the problem AND (2) show that the solution is right?

You should answer "yes" to this question only if a participant can do BOTH steps. Otherwise, answer "no."

If there is no well-defined "correct" or "best" answer to begin with, the answer to this question is always "no."

Finally, remember that this question is about whether it is possible to come up with ANY "best" solution and prove that it is right; if the proof is so complicated that others would not recognize it, you should answer "yes" to this question, because a proof of a correct solution still exists.

No, you CANNOT BOTH solve the problem and show the solution is right.

Yes, you can BOTH solve the problem and show the solution is right.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

Here is a reminder of the task:

Shopping plan

1. Set-Up

Each participant has a grocery list and a map that shows distances and times between each grocery store, as well as a list of potential items to purchase and how many points they are worth.

2. Objective / Goal

Participants have to plan a shopping trip as though they were all residents of the same house sharing the same car. The goal is to purchase as many high-quality items as possible in a fixed amount of time, and to get the highest number of points by considering tradeoffs between price, quality, and driving time. Participants gain points for every item that they are able to plan to purchase. Getting anything less than the maximum number of points will lead to partial credit.

QUESTION 3 (of 16). Can the outcome of this task be described in win/lose terms?

Elaboration: This question captures whether a task is a "competition" or "battle," in which your outcome is relative to someone or something else (e.g., another player, team, or an AI agent).

A battle is "where the focus is on conquest of an opponent and winner-take-all distribution of payoffs," and a competition is "where there is a lot of emphasis on

standards of performance excellence over and above the reckoning of winners and losers." Competitions between sports teams fit this category (e.g., basketball, soccer, ice hockey), but so too do military and street gang battles (e.g., the spoils of war go to the victor no matter how honorably the vanquished may have fought).

Answer 'yes' ONLY if you "win" or "lose" relative to somebody else who is playing the game. If someone performs poorly in the task, but they are not being compared to another player, you should answer 'no.' For example, if you are being asked to come up with as many ideas as possible, you may think that you "lose" if you generate zero ideas or that you "win" because you generated many ideas. However, you should still answer 'no,' because this question is asking about "winning" and "losing" in relative terms. A task in which you need to come up with more ideas than another player, on the other hand, would be a 'yes,' since you would win or lose relative to that other person.

No, the task is NOT a "competition" or "battle."

Yes, the task is a "competition" or "battle."

Not applicable or not answerable based on the task description (Please Elaborate Below.)

Here is a reminder of the task:

Shopping plan

1. Set-Up

Each participant has a grocery list and a map that shows distances and times between each grocery store, as well as a list of potential items to purchase and how many points they are worth.

2. Objective / Goal

Participants have to plan a shopping trip as though they were all residents of the same house sharing the same car. The goal is to purchase as many high-quality items as possible in a fixed amount of time, and to get the highest number of points by considering tradeoffs between price, quality, and driving time.

Participants gain points for every item that they are able to plan to purchase.

Getting anything less than the maximum number of points will lead to partial credit.

QUESTION 4 (of 16). Is the goal (or one of the goals) of this task to try to achieve doing something as much as possible, as many as possible, or as quickly as possible?

Elaboration: Sometimes the goal that is to be achieved entails doing as much as possible of something, or doing it as rapidly as possible. Thus, if an individual or group is asked to exert a maximum force on a rope, a strong pull is regarded as a more successful performance than a weak pull. If a team of mountain climbers is asked to ascend a cliff as rapidly as possible, maximum speed is the criterion against which performance is evaluated. Look for tasks asking participants to score the most number of points, get the most utility, for everyone to generate as many ideas as they can, and so on.

No, one of the goals is NOT about working as much/many/quickly as possible.

Yes, one of the goals is about working as much/many/quickly as possible.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

Here is a reminder of the task:

Shopping plan

1. Set-Up

Each participant has a grocery list and a map that shows distances and times between each grocery store, as well as a list of potential items to purchase and how many points they are worth.

2. Objective / Goal

Participants have to plan a shopping trip as though they were all residents of the same house sharing the same car. The goal is to purchase as many high-quality items as possible in a fixed amount of time, and to get the highest number of points by considering tradeoffs between price, quality, and driving time. Participants gain points for every item that they are able to plan to purchase. Getting anything less than the maximum number of points will lead to partial credit.

QUESTION 5 (of 16). If someone who is able to solve the problem explains their answer, would others recognize it as correct without contest?

Elaboration: This question is about whether participants who are not themselves able to solve the problem have sufficient knowledge of the system to recognize and accept a correct solution if it is proposed by someone else. In other words, are you informed enough to know the right answer when you see it?

If there is no well-defined "correct" answer to begin with, the answer to this question is always "no."

No, you CANNOT know the right answer when you see it.

Yes, you can know the right answer when you see it.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

Here is a reminder of the task:

Shopping plan

1. Set-Up

Each participant has a grocery list and a map that shows distances and times between each grocery store, as well as a list of potential items to purchase and

how many points they are worth.

2. Objective / Goal

Participants have to plan a shopping trip as though they were all residents of the same house sharing the same car. The goal is to purchase as many high-quality items as possible in a fixed amount of time, and to get the highest number of points by considering tradeoffs between price, quality, and driving time. Participants gain points for every item that they are able to plan to purchase. Getting anything less than the maximum number of points will lead to partial credit.

QUESTION 6 (of 16). Does completing this task require participants to evaluate tradeoffs --- conflicting possible solutions or conflicting pieces of information?

Elaboration: Some tasks provide a variety of inputs to help complete the task, but these inputs don't always agree or point to the same solution, and participants must evaluate or navigate tradeoffs in executing the task.

For example, a task involving writing an ad for a product may provide individuals with a variety of product features, customer reviews, and market information that may not all point to the same answer or angle for the ad. Similarly, a problem asking you to devise a business plan may require you to manage conflicting pieces of information, such as inconsistent needs for different consumer segments and tradeoffs between different courses of action.

No, there are NOT conflicting pieces of information or tradeoffs.

Yes, there are conflicting pieces of information or tradeoffs.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

Here is a reminder of the task:

Shopping plan

1. Set-Up

Each participant has a grocery list and a map that shows distances and times between each grocery store, as well as a list of potential items to purchase and how many points they are worth.

2. Objective / Goal

Participants have to plan a shopping trip as though they were all residents of the same house sharing the same car. The goal is to purchase as many high-quality items as possible in a fixed amount of time, and to get the highest number of points by considering tradeoffs between price, quality, and driving time. Participants gain points for every item that they are able to plan to purchase. Getting anything less than the maximum number of points will lead to partial credit.

QUESTION 7 (of 16). Is there an objectively correct solution to this task that can be calculated or selected?

Elaboration: If you answer 'yes' to this question, the task should have "a demonstrable right answer, and the group task is to invent/select/compute that correct answer." This right answer can be found in multiple ways:

- By using intuition
- By relying on general norms or accepted ideas
- By using logic or applying rules
- By consulting an expert (for example, the answer to, "is climate change real?" can be determined by an expert.)

If you answer 'no' to this question, "there is not a demonstrably correct answer, and ... the group's task is to select, by some consensus, a preferred alternative."

The alternatives might involve:

- Figuring out consensus on cultural values or moral courses of action
- Figuring out consensus by sharing relevant information

No, there is NOT an objective answer.

Yes, there is an objective answer.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

Here is a reminder of the task:

Shopping plan

1. Set-Up

Each participant has a grocery list and a map that shows distances and times between each grocery store, as well as a list of potential items to purchase and how many points they are worth.

2. Objective / Goal

Participants have to plan a shopping trip as though they were all residents of the same house sharing the same car. The goal is to purchase as many high-quality items as possible in a fixed amount of time, and to get the highest number of points by considering tradeoffs between price, quality, and driving time. Participants gain points for every item that they are able to plan to purchase. Getting anything less than the maximum number of points will lead to partial credit.

QUESTION 8 (of 16). When doing this task, will the participants have any uncertainty about whether their method or solution will lead to the desired outcome?

Elaboration: This question is asking whether a player knows for sure that their strategy or answer is right BEFORE the activity ends, or whether they have to wait until AFTER to find out. Answer 'yes' if participants submitting their answers may feel unsure that their submission or approach is correct, with no way to "check" their answer until the task is done.

Here are tasks where you would want to answer 'yes:'

- Tasks where you are playing against an opponent in a dynamic game that is not determined until the end. For a soccer game, you can come in with a "game plan" and a strategy and play your best, but that doesn't guarantee that you win. You have to play the whole game to find out who wins. Even if you're 6 points ahead, there's still a chance that the other team makes a comeback, so you do not know for sure that you won until the end. (The same goes for other types of games, like video games.)
- Tasks where there's a mathematical limit or objectively correct answer, which does not get revealed until the end. An example is guessing the number of jellybeans in a jar. You can try your best to guess, and you can even be fairly confident that you're right, but until the final answer is revealed, there is some uncertainty about whether your answer is "correct." Similarly, a task where you try to buy the most number of items with \$100 has a mathematical maximum; however, you may not be sure that you found that maximum until the end.

On the other hand, examples of 'no' are:

- Tasks where you can check your work easily. If you solve an algebra problem, you can solve it using a well-defined series of rules and steps. And you will know your answer is right, because you can substitute your solution for X back into the equation to confirm that it works.
- Tasks where you get information, BEFORE the game is over, that you already won. For example, if you need to earn at least 200 points and you can see your score update until it reaches 200, answer 'no.' Another example is if you need to move a small object from Point A to Point B --- you can just carry it out and confirm (by looking at its new location) that it worked.
- Tasks where the goal is to just do your best. If the task is to generate as many ideas as possible, there is no mathematical limit to how many you can come up with. As long as you do your best, you can be certain that you have

an acceptable answer. (On the other hand, if the goal is to generate more ideas than another player, then you would be uncertain.)

No, participants do NOT feel unsure. They can know that their answer is right during the task.

Yes, participants do feel unsure. They don't know how well they did until the end.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

The next 8 questions will be about the **Graph Coloring Task**. Please read about this task below. As before, we will show you a reminder of the task on every page.

The first question for this task appears on the next page.

Graph coloring task

1. Set-Up

Participants are given one or more “graph coloring problems”. These are pictures of different topology/network graphs.

For context, these look like many different nodes (points), which are connected by various edges. Imagine, for example, a picture of a social network: each person in the network is a “node,” and each of their friends is an “edge.”

In this case, all participants have access to and can interact with a shared graph. Participants will also be given a limited set of colors, with a larger number of colors than the minimum required to complete the task successfully.

2. Objective / Goal

The goal is to use the colors given to “color” the entire network without conflicts. This means that participants need to assign each node in the graph a color, ensuring that no two connected nodes share the same color. Using our social

network example, if Person A and Person B are friends (which means they are connected), they cannot be assigned the same color. Participants must successfully color all nodes and resolve all conflicts within a specified amount of time, or receive no credit.

Here is a reminder of the task:

Graph coloring task

1. Set-Up

Participants are given one or more “graph coloring problems”. These are pictures of different topology/network graphs.

For context, these look like many different nodes (points), which are connected by various edges. Imagine, for example, a picture of a social network: each person in the network is a “node,” and each of their friends is an “edge.”

In this case, all participants have access to and can interact with a shared graph. Participants will also be given a limited set of colors, with a larger number of colors than the minimum required to complete the task successfully.

2. Objective / Goal

The goal is to use the colors given to “color” the entire network without conflicts. This means that participants need to assign each node in the graph a color, ensuring that no two connected nodes share the same color. Using our social network example, if Person A and Person B are friends (which means they are connected), they cannot be assigned the same color. Participants must successfully color all nodes and resolve all conflicts within a specified amount of time, or receive no credit.

QUESTION 9 (of 16). Is there only one method for achieving the task, as opposed to many alternatives for task completion?

Elaboration: Is there only one right way of solving the problem?

Answer 'no' if there is more than one possible course of action or process to attain the group's goal.

Answer 'yes' if there is only ONE process or action that will lead to the correct answer or achieve the goal.

As an example, answer 'yes' if the instructions specify exact steps, e.g., "first you write all the 3-letter words, then the 4-letter words, then the 5-letter words," or if there's really only one way to do the task (e.g., perform long division). Actions might also be limited by the environment: participants may be working on an electronic system that restricts their communication or the steps that they are allowed to take.

However, answer 'no' if the task is open-ended in terms of course of action: e.g., if participants are simply asked to "come up with as solution" or "give as many ideas as possible" without specifying how they should achieve the goal.

No, there are MANY ways to achieve this task.

Yes, there is only ONE way to achieve this task.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

Here is a reminder of the task:

Graph coloring task

1. Set-Up

Participants are given one or more "graph coloring problems". These are pictures of different topology/network graphs.

For context, these look like many different nodes (points), which are connected by various edges. Imagine, for example, a picture of a social network: each person in the network is a "node," and each of their friends is an "edge."

In this case, all participants have access to and can interact with a shared graph. Participants will also be given a limited set of colors, with a larger number of colors than the minimum required to complete the task successfully.

2. Objective / Goal

The goal is to use the colors given to "color" the entire network without conflicts. This means that participants need to assign each node in the graph a color, ensuring that no two connected nodes share the same color. Using our social network example, if Person A and Person B are friends (which means they are connected), they cannot be assigned the same color. Participants must successfully color all nodes and resolve all conflicts within a specified amount of time, or receive no credit.

QUESTION 10 (of 16). Is the solution to the question obvious as soon as it is proposed --- for example, once people see the "trick," they know how to solve it?

Elaboration: Some questions have a special trick in them, such that, if you know the trick, the question is easy, but if you do not know the trick, the question may be quite difficult. An example of this is the Sphinx's Riddle (e.g., "what has 4 legs in the morning, 2 in the afternoon, and 3 in the evening?"). This puzzle is confusing if you have never heard it, but if you have heard the answer (a human -- since babies crawl, adults walk, and the elderly use canes), you'll never be confused again.

In other words, does this task cause a "Eureka!" or an "Oh, I get it!" moment when the answer is explained?

If there is no trick, answer 'no' to this question.

No, there is NOT an "aha" moment that makes this task easy once you see the trick.

Yes, there is an "aha" moment that makes this task easy once you see the trick.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

Here is a reminder of the task:

Graph coloring task

1. Set-Up

Participants are given one or more “graph coloring problems”. These are pictures of different topology/network graphs.

For context, these look like many different nodes (points), which are connected by various edges. Imagine, for example, a picture of a social network: each person in the network is a “node,” and each of their friends is an “edge.”

In this case, all participants have access to and can interact with a shared graph. Participants will also be given a limited set of colors, with a larger number of colors than the minimum required to complete the task successfully.

2. Objective / Goal

The goal is to use the colors given to “color” the entire network without conflicts. This means that participants need to assign each node in the graph a color, ensuring that no two connected nodes share the same color. Using our social network example, if Person A and Person B are friends (which means they are connected), they cannot be assigned the same color. Participants must successfully color all nodes and resolve all conflicts within a specified amount of time, or receive no credit.

QUESTION 11 (of 16). Is this a "generation" or "brainstorming" task? In other words, is one of the main purpose(s) of this task to produce a number of ideas or examples, without any particular action associated with them?

Elaboration: These are tasks based around generating ideas, examples, or concepts.

The examples being generated can be either abstract (generating words and colors) or concrete (generating ideas for how to spend \$10,000 or ways to use a paperclip). If the ideas are concrete, they should NOT suggest a specific plan or course of action. This question is therefore different from the previous question; the participants are NOT generating a plan.

Note that you should only answer "yes" to this question if one of the main outcomes of this task is to generate such ideas. Many tasks require people to discuss ideas (for example, a jury task), or come up with creative ways to solve a problem (for example, a math problem), but the primary goal of such tasks is not just to generate the ideas.

No, this is NOT a "generation" or "brainstorming" task.

Yes, this is a "generation" or "brainstorming" task.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

Here is a reminder of the task:

Graph coloring task

1. Set-Up

Participants are given one or more "graph coloring problems". These are pictures of different topology/network graphs.

For context, these look like many different nodes (points), which are connected by various edges. Imagine, for example, a picture of a social network: each person in the network is a "node," and each of their friends is an "edge."

In this case, all participants have access to and can interact with a shared graph. Participants will also be given a limited set of colors, with a larger number of colors than the minimum required to complete the task successfully.

2. Objective / Goal

The goal is to use the colors given to "color" the entire network without conflicts. This means that participants need to assign each node in the graph a color, ensuring that no two connected nodes share the same color. Using our social network example, if Person A and Person B are friends (which means they are connected), they cannot be assigned the same color. Participants must successfully color all nodes and resolve all conflicts within a specified amount of time, or receive no credit.

QUESTION 12 (of 16). Is the goal (or one of the goals) of this task to try to achieve a precise outcome?

Elaboration: Here, there is a specific standard to meet, but precision matters.

Look for tasks with a specific, most preferred outcome. If the task is to estimate the temperature of a room, the goal is to agree with the value indicated by the thermometer. Another example is moving exactly 10 objects --- no more, no less --- or to exactly reproduce something (such as a task where people have to exactly copy a work of art).

Also look out for concepts like "exactly," "precisely," "optimal," or "best." For example, if you are asked to generate exactly 10 ideas, generating both 9 ideas and 11 ideas would be considered a failure. Another clue might be terms like "constraints," "requirements," or "rules" that someone has to follow, as long as there is a specific best answer within the rule set. For example, if you want to buy the most number of items using a budget of \$100, the "constraint" of \$100 also serves as a precise goal. That is, the closer you get to \$100 (the more precise you are), the better you do at the task.

No, one of the goals is NOT to achieve a precise outcome.

Yes, one of the goals is to achieve a precise outcome.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

Here is a reminder of the task:

Graph coloring task

1. Set-Up

Participants are given one or more “graph coloring problems”. These are pictures of different topology/network graphs.

For context, these look like many different nodes (points), which are connected by various edges. Imagine, for example, a picture of a social network: each person in the network is a “node,” and each of their friends is an “edge.”

In this case, all participants have access to and can interact with a shared graph. Participants will also be given a limited set of colors, with a larger number of colors than the minimum required to complete the task successfully.

2. Objective / Goal

The goal is to use the colors given to “color” the entire network without conflicts. This means that participants need to assign each node in the graph a color, ensuring that no two connected nodes share the same color. Using our social network example, if Person A and Person B are friends (which means they are connected), they cannot be assigned the same color. Participants must successfully color all nodes and resolve all conflicts within a specified amount of time, or receive no credit.

QUESTION 13 (of 16). Can acceptable solutions to this task be demonstrated or verified to be correct (e.g., by an expert or third-party)?

Elaboration: This item refers to the "degree to which acceptable solutions can be demonstrated to be correct," via logic and rules as opposed to having a general consensus.

For example, the solution to a math problem can be verified via the rules of algebra. In a task where participants are asked to buy the items with the best value, one can also use logic to list out every possible combination of items and show that their solution is the best one. Another type of demonstration could involve showing facts: for example, if the task is to estimate how many people live in the United States, the "ground truth" is the statistic from the U.S. Census.

Examples of tasks where you should answer 'no' include answering the question, "should we ban all guns?" --- this cannot be verified via logic and rules. Evaluating

whether something is "creative," as well as other subjective judgements, is also not demonstrable.

No, you CANNOT verify a solution to be correct.

Yes, you CAN verify a solution to be correct.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

Here is a reminder of the task:

Graph coloring task

1. Set-Up

Participants are given one or more "graph coloring problems". These are pictures of different topology/network graphs.

For context, these look like many different nodes (points), which are connected by various edges. Imagine, for example, a picture of a social network: each person in the network is a "node," and each of their friends is an "edge."

In this case, all participants have access to and can interact with a shared graph. Participants will also be given a limited set of colors, with a larger number of colors than the minimum required to complete the task successfully.

2. Objective / Goal

The goal is to use the colors given to "color" the entire network without conflicts. This means that participants need to assign each node in the graph a color, ensuring that no two connected nodes share the same color. Using our social network example, if Person A and Person B are friends (which means they are connected), they cannot be assigned the same color. Participants must successfully color all nodes and resolve all conflicts within a specified amount of time, or receive no credit.

QUESTION 14 (of 16). Does the task have all-or-nothing outcome? In other words, is it sufficient to just meet a particular standard?

Elaboration: These tasks are about doing just enough work to meet some threshold or standard, but once you reach that standard, precision does not matter.

As an example, a pass-fail class (e.g., "anything above 70% is a pass") is all-or-nothing. On the other hand, a letter grade (A+ for 100%, A for 90, and so on) is not. Another example is a task in which people simply need to move objects from Point A to Point B: if they meet the outcome, they succeed. Even if they moved the object part of the way there, they don't get any credit for their effort. Moving it an extra 10 feet does not get them any brownie points.

Counter-examples include games in which you try to earn as many points as possible, and you end up with a numeric score instead of simply a success/fail. For example, trying to generate as many ideas as possible, or solve as many problems as possible, would NOT be all-or-nothing tasks. If participants moving an object get credit for every foot that they manage to move it, then it's NOT all-or-nothing.

No, this task does NOT have an all-or-nothing outcome.

Yes, this task has an all-or-nothing outcome.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

Here is a reminder of the task:

Graph coloring task

1. Set-Up

Participants are given one or more "graph coloring problems". These are pictures of different topology/network graphs.

For context, these look like many different nodes (points), which are connected by various edges. Imagine, for example, a picture of a social network: each person in the network is a "node," and each of their friends is an "edge."

In this case, all participants have access to and can interact with a shared graph. Participants will also be given a limited set of colors, with a larger number of colors than the minimum required to complete the task successfully.

2. Objective / Goal

The goal is to use the colors given to "color" the entire network without conflicts. This means that participants need to assign each node in the graph a color, ensuring that no two connected nodes share the same color. Using our social network example, if Person A and Person B are friends (which means they are connected), they cannot be assigned the same color. Participants must successfully color all nodes and resolve all conflicts within a specified amount of time, or receive no credit.

QUESTION 15 (of 16). Can this task be written as a "formal model" that an algorithm could solve?

Elaboration: Some tasks can be written as a formal model or math problem, expressed with rules and syntax that the problem-solvers share.

One way to think about this question is to ask yourself, could a robot or algorithm do this task?

Here are examples of a 'yes' answer:

1. A specific set of rules and outcomes: Robots are really good at following rules. Answer 'yes' if you can input the rules and the desired outcome, and have a robot follow predetermined steps to get the result. For example, you can tell the robot, "buy as many items as you can with a budget of \$100." (In fact, online shopping websites have exactly this tool!)
2. You can write it like a math problem: Robots are really good at doing math. Something like, "find the best teams under the constraints," combined with a list of how much "utility" people get from being put together, can easily be put into a computer and solved. Similarly, if you're trying to find the shortest path from Point A to Point B, you can think of this in terms of geometry.

3. You can break the problem down into small units of meaning: Examples of these units include the alphabet (26 letters), colors (which can be broken down into units of Red, Blue, and Green), or coordinates on a graph (X,Y). An algorithm could use these units of meaning to solve the task. For example, when generating words, the letters are the units. You can imagine an algorithm in which you use "brute force" to find every possible combination of letters that makes a valid word. Another example is Wordle --- many people have built bots that consistently solve the puzzle!

Here are examples of a 'no' answer:

1. Creative ideas that cannot be generated just by following basic rules. Whereas you can generate words just by using the rule of trying out all the different letters in the alphabet, you can't generate stories in the same way. There are no "rules" for what makes a good story. (We ask you not to think about advanced models like GPT-3 for the purposes of this question.)
2. Subjective tasks that require a judgment call. Deciding whether or not we should ban all guns can't really be done by an algorithm: people have to debate these ideas.
3. Tasks relying on social dynamics. Anything focusing on relationships between participants should be labeled 'no.'

No, this is NOT a formal problem that an algorithm could solve.

Yes, this is a formal problem that an algorithm could solve.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

Here is a reminder of the task:

Graph coloring task

1. Set-Up

Participants are given one or more "graph coloring problems". These are pictures of different topology/network graphs.

For context, these look like many different nodes (points), which are connected by various edges. Imagine, for example, a picture of a social network: each person in the network is a "node," and each of their friends is an "edge." In this case, all participants have access to and can interact with a shared graph. Participants will also be given a limited set of colors, with a larger number of colors than the minimum required to complete the task successfully.

2. Objective / Goal

The goal is to use the colors given to "color" the entire network without conflicts. This means that participants need to assign each node in the graph a color, ensuring that no two connected nodes share the same color. Using our social network example, if Person A and Person B are friends (which means they are connected), they cannot be assigned the same color. Participants must successfully color all nodes and resolve all conflicts within a specified amount of time, or receive no credit.

QUESTION 16 (of 16). Is there enough information in the problem to find a valid solution?

Elaboration: This item refers to whether there is sufficient information to obtain a solution within the system. In other words, this question is about whether it is possible --- using all of the information provided to the participants --- to get an valid or acceptable answer.

By "the system," we mean the entire set of stimuli given to the participant(s) for solving the problem. If the person is rating images, it is the set of images and the rating scale/survey; if it is a game, then it is the entire self-contained "system" of the game and the rules for playing it. If it is an optimization problem, it's the set of object(s) that need to be optimized and the list of constraints.

Examples of 'no' answers:

1. The only way to get a valid solution is outside of the "system" --- e.g., looking the answer up, asking someone else. For example, a trivia quiz (where you have to look up the right answer) would be a 'no' for this question.

2. There is no right answer, or the question is fundamentally unsolvable (e.g., it's a trick question).
3. There's not enough information to solve the question. In math, for example, you need two equations to solve for two unknown variables. If you don't have enough information, you can't do the problem.

Examples of 'yes' answers:

1. Any question where you can get a right answer from the information given is a 'yes.' This is true even if you're not sure of your answer. For example, in a multiple-choice quiz, you may not be sure that your answer is correct. But if one of the choices was the right answer, and the problem gave enough information for you to select it, then answer 'yes.'
2. If there are multiple valid answers in response to the problem (e.g., a creative writing task where any valid story is accepted), the answer should be 'yes.'

No, there is NOT enough information to find a solution.

Yes, there is enough information to find a solution.

Not applicable or not answerable based on the task description (Please Elaborate Below.)

Feedback for Us

You've reached the end of the survey; thank you for participating and for all your effort up to this point!

Before submitting, we have some final questions about your answers to this survey.

Did you answer questions in this survey randomly, as opposed to thinking carefully about your response, at any point?

(Please note that your response to this question does NOT affect your payment.)

No

Yes

Do you think there is any reason that we should NOT use your responses for this survey?

(Please note that your response to this question does NOT affect your payment.)

No

Yes

Finally, **do you have any feedback or comments about your experience?** We will use your comments to improve this survey for the future.

