Judgment and Decision Making (2025), **vol. 20**:e32 1–32
doi:10.1017/jdm.2025.10008

EMPIRICAL ARTICLE

Development and initial validation of a situational judgment test for the measurement of actively open-minded thinking

Nikola Erceg, Andrija Vrhovnik, Zvonimir Galić, and Mitja Ružojčić

Faculty of Humanities and Social Sciences, University of Zagreb, Zagreb, Croatia

Corresponding author: Nikola Erceg; Email: nerceg@ffzg.hr

Received: 11 June 2024; **Revised:** 25 July 2025; **Accepted:** 25 July 2025

Keywords: actively open-minded thinking; situational judgment test; validation

Abstract

Existing measures of Actively Open-Minded Thinking (AOT) primarily assess the acceptance of rational thinking norms and standards, rather than actual thinking and resulting behavior. These scales can be susceptible to impression management, often yield inflated scores, and may not accurately capture how individuals think in real-life contexts. To address these limitations, we developed and validated a novel Situational Judgment Test for Actively Open-Minded Thinking (AOT-SJT), designed to assess behavioral tendencies related to AOT in realistic scenarios. AOT is conceptualized as the disposition to consider alternative viewpoints, seek disconfirming evidence, and revise beliefs in light of new information. Across 4 studies, we constructed and refined the AOT-SJT using scenarios that simulate everyday decision-making. In Study 1, we tested initial items among Croatian participants, resulting in a 13-item measure with solid psychometric properties. Study 2 confirmed the test's convergent validity with cognitive and personality constructs and its predictive power for different forms of rational thinking. In Study 3, new items were introduced to enhance construct coverage, particularly around evidence search direction. Study 4 extended validation to an English-speaking sample, supporting cross-linguistic applicability, although effect sizes related to convergent validity were somewhat lower than before. Findings across studies show that the AOT-SJT aligns with theoretical expectations, demonstrates solid convergent validity with existing AOT scales, and effectively distinguishes levels of open-mindedness. By measuring behavioral intentions rather than standards acceptance, the AOT-SJT offers an externally valid assessment of AOT.

1. Introduction

We make hundreds of decisions every day, in different spheres of life and of different importance (Milosavljevic et al., 2011). Not all our decisions are correct or completely based on rational thinking. We make most decisions without intensive thinking, using mental shortcuts that speed up the thinking and decision-making process (Ellis, 2018). In many situations, this strategy is sufficiently accurate and successfully guides our everyday behavior (Gigerenzer and Gaissmaier, 2011), but despite this, these shortcuts or heuristics frequently can result in errors in thinking and cognitive biases that manifest themselves as bad decisions, especially in complex decision situations where thorough thinking is needed (e.g., Nutt, 1999; Sibony, 2020).

Baron (1991) introduced the concept of *actively open-minded thinking (AOT)* as an ideal way of thinking that would counteract many pervasive cognitive biases and that can be used as a norm for the evaluation of the quality of thinking. AOT is defined by Baron as the tendency to impartially judge various possibilities despite the fact that some of them oppose the one we initially favored.

Baron based AOT on the previously presented theory of good thinking in his book *Rationality and Intelligence* (Baron, 1985). There, he set a framework for evaluating thinking based on the search for possibilities, evidence, and goals while drawing conclusions from them. Baron (1985) defined good thinking as an optimal search for possibilities, evidence, and goals that is unbiased toward all evidence and possibilities. Moreover, AOT as an ideal does not only represent openness to the reasons why we are wrong, but also an active search for them. This means that the amount of search is proportional to the importance of the question and is open to possibilities different from the one we initially favored, while the confidence in the decision reflects the amount and quality of the thinking done.

AOT can also be viewed as a prescriptive model for avoiding common reasoning failures. Baron (1991) emphasized that poor thinking is typically characterized by insufficient search and a bias toward favored conclusions, whereas Baron (2019) highlighted both the amount and the direction of cognitive search as critical parameters. He argued that good thinking requires a level of search effort proportional to the importance of the question and that people generally underthink when deeper consideration is warranted. In contrast, good thinking—AOT—consists of:

a. thorough search, scaled to the importance of the issue,
b. fairness to alternative possibilities, especially those that counter one's initial preference, and
c. confidence levels calibrated to the amount and quality of thinking undertaken (Baron, 2023).

Thus, when measuring AOT, it is sensible to assess how people differ across these 3 aspects of good thinking. However, a recent perspective on AOT (Baron, 2024) emphasizes 2 key components of individual differences: (a) susceptibility to myside bias (related to openness and fairness toward alternative possibilities) and (b) aversion to uncertainty, which often manifests as overconfidence. From this perspective, a valid AOT measure should capture variation in these 2 domains—precisely what Baron's (2024) revised measure aims to do.

Nevertheless, in our current work, we adopt a more expansive view of individual differences in AOT. In addition to the 2 components highlighted by Baron (2024), we also consider a third: the quantity of search, or the cognitive effort devoted to thinking and evidence gathering, where people typically err by doing too little rather than too much. Thus, in this work, when conceptualizing AOT and its measurement, we consider 3 interrelated components:

a. quantity of search—engaging in sufficient thinking proportional to the importance of the decision,
b. direction of search—actively seeking counterarguments to challenge initial beliefs, and
c. confidence calibration—avoiding overconfidence by aligning certainty with the depth and quality of one's reasoning.

In sum, AOT can serve as a norm for evaluating one's own opinion, as a set of dispositions that describe a way of thinking that is consistent with the norm, and as a standard for evaluating other people's opinions (Baron, 2019). It is a prescriptive model that does not focus on the success or failure of the decision made, but on the thinking process behind it. The reasons for inadequate thinking are typically found in (a) insufficient search for possibilities and goals, especially those different from ones we currently favor, and (b) overconfidence in the chosen option, which occurs when high confidence is not justified by the quality of thinking and is most often accompanied by confirmation bias (Baron et al., 2022). In essence, AOT is a style of thinking that leads to less biases and better judgments/decisions by counteracting these reasoning errors.

1.1. AOT and other indicators of rational thinking

In line with AOT being a model of rational thinking, it has been shown to be a good predictor of performance in critical thinking, rationality, and decision-making tasks (Janssen et al., 2020). It is also positively correlated with cognitive reflection (Frederick, 2005) and other indicators of analytical thinking, such as the need for cognition (NFC; Haran et al., 2013) and the rational cognitive style (Rational-Experiential Inventory—short [REI]; MacLaren et al., 2012). Moreover, AOT is positively

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

correlated with various measures of decision task performance (Baron, 2019), such as the persistence in search for information and estimation of precision, and negatively correlated with the overconfidence bias (Haran et al., 2013). Overconfidence bias manifests itself through the tendency of people to overestimate the accuracy of their answers or their success in a given task. A good example is the cross-cultural research by Stankov and Lee (2014), which found that, on average, participants exhibited an overconfidence bias of 20%, obtained by the difference between the average confidence in the accuracy of the performance and the average accuracy of the actual performance.

In addition to success in tasks that require thinking and rationality, AOT has been shown to be correlated with different beliefs. It is positively correlated with the acceptance of the theory of evolution (Athanasiou and Papadopoulou, 2012) and negatively correlated with political conservatism (Baron et al., 2023), supernatural beliefs (Svedholm-Häkkinen and Lindeman, 2018), belief in conspiracy theories (Janssen et al., 2020) and conspiratorial mentality, a general tendency to interpret events through conspiratorial explanations and beliefs (Pennycook et al., 2020; Swami et al., 2011, 2014; Zajenkowski et al., 2022). Finally, AOT has also been shown to be correlated with some behaviors. For example, it was negatively correlated with pathological gambling (MacLaren et al., 2012) and positively correlated with the tendency to recognize misinformation (Roozenbeek et al., 2022).

1.2. Measuring actively open-minded thinking

Based on Baron's idea of AOT assessment through the measurement of attitudes that refer to certain standards of thinking, Stanovich and West (1997, 2007) developed the AOT scale. It is the most used instrument for assessing AOT, which measures the tendency for AOT in the form of attitudes toward what constitutes good thinking. Examples of some items are 'People should take into consideration evidence that goes against conclusions they favor' and 'Changing your mind is a sign of weakness' (reverse scoring). Over the years, numerous versions of this scale have been developed. Janssen et al. (2020), in their review of research in the period from 2007 to 2019, count as many as 36 versions of the AOT scale, with the 41 items being the most common one. Since all the abovementioned relationships between AOT and other constructs were established by applying the AOT scale, it appears to be a valid measure of this disposition.

However, one of the main problems with the scale is that the AOT scale is only a measure of attitudes toward standards of thinking or acceptance of those standards (Baron, 2018). This offers an advantage that ordinary self-report scales lack: the standards assessed can be applied to the assessment of other people's trustworthiness as sources of information. Still, the AOT scale measures attitudes toward a way of thinking, but not how someone actually thinks. Furthermore, the AOT scale is susceptible to impression management. It is reasonable to assume that respondents might agree with the thinking according to the standards of AOT to present themselves in a more favorable light. Perhaps because of this, the scores on the AOT scale are often high (Baron, 2018), which negatively affects the reliability and validity of the instrument. Baron (2018) points out the need to create more realistic and ecologically valid measures that would better encompass the construct of AOT and make a clearer distinction against measures of similar concepts. Indeed, Baron (2024) introduced a measure designed to assess 2 key components of AOT—myside bias and uncertainty aversion—across a range of situations. An additional goal of this measure was to examine whether AOT is generalizable across contexts; that is, whether individuals who endorse AOT actually apply it consistently in different situations. We believe that a situational judgment test, a measurement system with a long research history in industrial/organizational psychology, might be an approach that could both improve the deficiencies of the AOT scale and be used as an additional way of testing the generality of AOT across a wide range of reasoning and decision-making situations.

1.3. Situational judgment tests

Situational judgment tests (SJTs) typically contain scenarios that describe a relevant situation and a list of several possible (behavioral) responses to that situation (Oostrom et al., 2019). A respondent's task

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

is to choose the answer that most closely describes how (s)he would or should behave in the given situation. The basic logic behind SJTs lies in the fact that the chosen response to the described situation in the scenario predicts the behavior in a similar real-life situation in the future because it captures respondents' behavioral intentions (Lievens and De Soete, 2015).

Most SJTs were developed within organizational psychology for personnel selection purposes. A meta-analysis by McDaniel et al. (2007) found that SJTs predict work performance even when the results are adjusted for the effects of cognitive ability, personality, and work experience. Compared with typical measures of individual characteristics that use Likert-type scales, the advantages of SJTs are higher predictive validity, lower risk of giving socially desirable answers, and greater item contextualization and realism (Olaru et al., 2019; Oostrom et al., 2019).

1.4. This research

The limitations of existing AOT measures prompted the undertaking of this study. We believe that a potential avenue is the development of an instrument based on the SJT paradigm that captures behavioral intentions related to AOT in realistic social situations, but that is, at the same time, grounded in the construct theory, encompassing all aspects of AOT. Therefore, this study focuses on the development and validation of a situational judgment test to measure AOT.

Unlike conventional Likert-type scales used in previous AOT measures, the SJT does not require participants to express their attitudes about a particular thinking style. In line with the mentioned characteristics of SJTs, we believe that responses to scenarios that describe everyday events reveal behavioral tendencies related to AOT, thus better reflecting how somebody would think and make decisions in real situations. Therefore, the overall score on the AOT-SJT should represent a more direct measure of participants' general inclination toward AOT in various simulated situations than measures that ask participants to express their attitudes toward particular thinking styles.

2. Overview of the present studies

In this article, we describe 4 studies across which we developed the new situational judgment test for AOT assessment (AOT-SJT). In developing AOT-SJT, we employed a theory-based approach in line with the previously described conceptualization of AOT with 3 key elements: the direction of search, the amount of search, and confidence in the decision made. To check the construct validity of the AOT-SJT, we examined its relationship with the latest short version of the AOT scale (Baron et al., 2022), as we assume that the AOT scale should, to a certain extent, tap into the same construct, but also many additional variables from its nomological network.

In Study 1, we developed and tested an initial version of AOT-SJT consisting of 20 scenarios with the goal of identifying the best candidate items for the final AOT-SJT measure. Study 1 resulted in a shorter, 13-item-long AOT-SJT whose psychometric characteristics and validity we investigated in Study 2. Study 3 was conducted with the goal of complementing the existing SJT items measuring the direction of search for evidence with additional items to strengthen that aspect of the measure and improve its psychometric properties. Finally, as the first 3 studies were done on Croatian participants, we conducted the fourth and final study to validate the full AOT-SJT on English-speaking participants.

3. Study 1

3.1. Construction of the AOT-SJT

In Study 1, we report on a study where we developed and validated the initial SJT-AOT version. To devise a large enough pool of situations appropriate for the expression of AOT tendencies and response options that indicate different levels of AOT responses, we engaged master-level psychology students

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

from the University of Zagreb whose task was to come up with scenarios and response options for that scenario. Students were enrolled in a course related to personnel selection with a specific emphasis on the usefulness and construction of situational judgment tests. A practical aspect of the course was the development of SJT, where students were engaged in the development of AOT-SJT. Prior to that, the students were provided with a comprehensive overview of the concept of AOT and its defining features and then asked to come up with real-life scenarios that provoke AOT-consistent or inconsistent responses. In addition to the scenarios, students also came up with 4 viable response options, where theoretically each option indicated a different level of AOT-consistent behavior. Each student was assigned one of the 3 defining aspects of AOT (quantity of search, direction of search, or overconfidence avoidance) and tasked to come up with a scenario and response options reflecting different levels of that specific AOT aspect. Students were instructed to come up with scenarios and responses from any sphere of life, as our intention was to cover the broadest possible range of situations in which a person can exhibit AOT-consistent behavior. This process resulted in 61 items in total.

The next step was choosing a smaller set of items for initial empirical validation of AOT-SJT. To do this, the authors first carefully read and graded the quality of each of the items devised by the students. When deciding on the quality of items, we looked for scenarios that were realistic, nontrivial, and allowed for the expression of AOT-consistent behavior. Furthermore, we wanted to sample a broad range of situations as when dealing with formative measures, only a heterogeneous and representative sample of situations ensures that the construct is measured comprehensively (Bledow and Frese, 2009). Finally, we favored situations that introduced some sort of challenge for acting in AOT-consistent ways (e.g., lack of time and disapproval from others). This ensured 2 things: (a) that choosing AOT-consistent behavior reflects a real belief that the AOT way of thinking and behaving is preferred and not just socially desirable low-cost behavior, and (b) that the responses indicating AOT-inconsistent behavior seem as a viable and reasonable option as opposed to being *a priori* discarded as wrong. Apart from the situations, we also assessed the quality of response options where our primary concern was that different response options actually reflect different levels of construct-consistent behavior, but also to check that some of the response options are not evidently inappropriate for a given situation.

This process resulted in 20 items that we then tested in Study 1, 6 assessing the quantity of the search aspect of AOT, 6 assessing the direction of search, and 8 assessing an overconfidence tendency. However, prior to empirical testing, we decided to somewhat change the response options. Specifically, our students were instructed to come up with 4 response options that indicate different levels of AOT-consistent behaviors, ranging from the most AOT-inconsistent behavior (i.e., indicating that a person behaves contrary to AOT principles) to the most AOT-consistent behavior (i.e., a person behaves completely in an AOT-consistent way). However, after going through all the scenarios and responses, we concluded that, in many cases, response options do not differentiate levels of AOT-consistent behavior. Thus, we decided to remove one response option and to proceed with only 3, but more differentiated response options.

Before describing the methods and results of Study 1's empirical investigation, we will explain the logic behind each response option on a sample item from each of the 3 aspects of AOT. Although the scenarios related to each of the AOT aspects were quite heterogeneous, the response options followed a similar logic consistently.

3.1.1. Quantity of search

The items designed to assess AOT-consistent behavior related to the quantity of search aim to capture an individual's behavioral tendency to invest more time and effort in gathering information and evidence when making decisions that are arguably important. This dimension focuses solely on the amount of search, rather than its direction—that is, it does not address whether the individual seeks information that supports or contradicts their initial beliefs.

It is important to acknowledge that more thinking or extended information search is not always more rational or desirable. In some cases—such as when the decision is trivial or unimportant—it may be unnecessary or even inefficient to engage in deeper cognitive effort. Similarly, when expert guidance is

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Table 1. *A sample item from SJT-AOT (quantity of search).*

You have been working in your field for 3 years at an organization where you were employed right after college. You are satisfied with your current job and relationships with colleagues, but occasionally encounter misunderstandings with your supervisor. You have received a job offer from another organization, and after successfully going through the selection process, you have been offered the same position with a 10% higher salary than your current one. They have asked for your response within 4 days. What should you do in this situation?
a. Considering it is a good opportunity and offers a higher salary, you will accept the job immediately. A quick response will demonstrate motivation and decisiveness. If you do not respond promptly, the company might hire someone else.
b. You will have a conversation with an acquaintance who works at the organization that offered you the job. You will ask about their experiences and seek advice. After this discussion, you will make a decision as soon as possible.
c. You will express gratitude for the offer and respond that you will need to think about it. Over the next few days, you will thoroughly research the other organization and try to gather information from their employees. You will consult with your colleagues and consider the pros and cons of your current job and the job offered before making a decision.

readily available and clearly more reliable than one's own judgment, relying on personal deliberation may actually lead to poorer outcomes. In such instances, excessive search may be counterproductive.

We carefully considered these issues when designing our scenarios. Each scenario was constructed to involve a nontrivial and moderately complex decision—one where a reasonable person would be expected to engage in some level of thoughtful deliberation. Importantly, none of the scenarios involved the possibility of consulting an expert whose advice would be clearly superior or objectively correct. In each case, the individual is placed in a situation where they must make the decision independently, based on their own information search.

Thus, all scenarios are designed to reflect plausible, real-life situations where more extensive evidence gathering is a rational and appropriate response. The behavioral options provided vary primarily in the amount of time and effort devoted to searching for relevant information, allowing us to assess individual differences in the tendency to engage in AOT-consistent thinking in the quantity domain.

However, we will stress two possible drawbacks of our items. First, we acknowledge that the judgment about whether the decision is important and nontrivial or not is a subjective one, and that possibly not all people will agree with us on the importance of these decisions. This might introduce an additional source of error in measurement as people might choose suboptimal responses not because they do not agree and follow AOT principles, but because they do not think the decision at hand is worth exhibiting them. An additional limitation of our current item pool is that it does not allow us to detect individuals who might depart from AOT by overthinking—that is, engaging in excessive deliberation even when the situation does not warrant it. We intentionally avoided including trivial or low-stakes decisions where such overthinking tendencies might become apparent. Our goal was to simplify the assessment by focusing on situations where more thinking is generally rational and beneficial. While this decision enhances clarity in interpreting results, it does come at the cost of reduced sensitivity to a less common, but still meaningful, form of AOT departure.

Having said that, an example of an item intended to capture the quantity of search is given in Table 1, and the complete set of items from the final AOT-SJT version is given in Table A1 in the Appendix.

The logic behind response options is that, given that leaving relatively good job for a new one is quite important decision, response 'c' reflects the probable behavior of a person who believes that AOT is appropriate way of thinking in such situations and who additionally would 'walk the talk', that

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Table 2. *A sample item from SJT-AOT (direction of search).*

After a routine check-up, the doctor approaches you with feedback. They inform you that you are generally in good health but have identified a potential health concern among the various other findings. While a relatively routine and straightforward surgical procedure could ensure that the issue doesn't become serious, the doctor assures you that it won't be necessary and that there is no need for concern. As you have never noticed any symptoms or had issues related to the concerns described, you also feel that excessive worry might not be warranted, but you still want to gather additional information. What would you do in this situation to make the best possible decision?
a. You will ask the doctor to provide you with a more detailed explanation of why they believe the surgical procedure is unnecessary. You are confident that this will give you a clearer understanding of what to do and help you feel more at ease.
b. You will request more information from the doctor about the specific health concern mentioned, and you will inquire about the experiences of individuals who have undergone the surgical procedure as well as those who have not.
c. You will ask the doctor to provide you with a more detailed description of this health concern, its symptoms, and long-term prognosis. In doing so, you will specifically ask them to outline the risks associated with not undergoing the surgery.

is, choose to behave in an AOT-consistent way despite some challenges for doing so. On the contrary, response option 'a' indicates that the person either does not see the situation as important, does not favor AOT in general, or does not believe that AOT-consistent behavior is appropriate in a given situation. The middle option is just that, somewhere in the middle between these 2 positions. Thus, the item is scored so that the lower number of points is assigned to response option 'a' (1 point), followed by response option 'b' (2 points), and finally by the 'best' option 'c' (3 points).

3.1.2. Direction of search

Items designed to assess behavioral tendency toward other-side information search are designed in a way that scenarios put the person in a decision-making situation that warrants additional search for information before deciding, while response options vary the direction of search. Therefore, a person prone to AOT should detect the danger of one-sided search and confirming already held position, and instead opt for searching for information and evidence that could counter his/her current position. An example of an item from this domain of AOT is given in Table 2.

In this example, a person is confronted with a situation that requires more investigation before making a decision. The most AOT-consistent response ('c') requires a person who (a) does not have any symptoms, (b) would probably want to avoid unpleasant and risky surgery, and (c) has a doctor that confirms that the surgery is not necessary to search for information and evidence that could prove him/her wrong, at the expense of personal discomfort and going against doctor's advice. Therefore, a person who really believes in the merits of AOT and does not hesitate to behave in an AOT-consistent way would most likely choose this response. Contrary to this, response option 'a' indicates a willingness to search for and listen only to arguments that confirm, and not endanger, one's prior position. Option 'b' is again in the middle between these 2, opting for additional search, but not quite in a disconfirming way. We would say that what distinguishes the 'c' response from the 'b' response across all items targeting the direction of search is that the 'c' option consistently reflects an active search for counterarguments—that is, a deliberate effort to challenge one's initial or preferred position, rather than leaving it to chance or relying on others to point out potential flaws. In line with this, option 'a' again is given 1 point, 'b' 2 points, and 'c' 3 points.

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Table 3. *A sample item from SJT-AOT (overconfidence).*

You are a red wine enthusiast. For many years, you have been educating yourself about it, and you are a favorite among your friends because you always know how to choose a good wine and describe it in a way others do not. You regularly consume it—you have a glass every day after lunch or dinner, and you are happy because you've read the studies that wine is good for the cardiovascular system. In a conversation with a friend, you learn about the latest research he read about on an internet portal that claims that even the smallest amount of red wine does not contribute to health. On the contrary, the research shows that all types of alcohol, including wine, are ultimately harmful. What should you do in this situation?
a. You will disregard this information. The contents of internet portals are often inaccurate, and the new research is likely questionable, given the conflicting results of all previous studies. You will explain to your friend that he should not trust such information.
b. You are willing to listen to the new research. While you believe in the accuracy of previous information, there is always a chance that previous studies were wrong. However, you still believe that wine can contribute to health in the way you consume it.
c. After hearing about the new research, since you are hearing about it for the first time, you are no longer as certain about the benefits of red wine. You know that newer research can uncover something that previous studies missed. You admit to your friend that there is a good chance he might be right.

3.1.3. Avoidance of overconfidence

The scenarios in items assessing the avoidance of overconfidence put a respondent in a situation in which (s)he holds a relatively strong opinion about some issue, but new information or evidence appears that should make a person revise his/her beliefs. Response options are then construed in a way that allows a person to double down and stick to his/her conviction, or to loosen up and revise his/her confidence in their own position.

One possible objection to our operationalization of overconfidence is that it may conflate it with rigidity—an excessive unwillingness to change one's mind. While Baron's (2024) recent conceptualization of AOT links overconfidence primarily to uncertainty aversion rather than rigidity, the two are often closely intertwined in behavior.

Several points are important to emphasize in this regard. First, although it is theoretically possible for someone to be rigid without being overconfident—for example, by acknowledging uncertainty yet still refusing to revise their stance—this is a peculiar position and, in our view, still indicative of low AOT. After all, a core tenet of AOT is the willingness to revise one's conclusions in response to relevant new information. Second, as a behaviorally based measure, our SJT necessarily captures observable indicators of overconfidence, and an unwillingness to reconsider one's views is among the most salient. Because overconfidence and rigidity often reinforce each other, we view resistance to changing one's mind as a valid and reliable behavioral manifestation of overconfidence.

Third, even existing measures of AOT standards acceptance frequently blur the line between overconfidence and rigidity (e.g., items like 'Changing your mind is a sign of weakness'), suggesting that this conceptual overlap is not unique to our approach but is inherent to how the construct is typically measured. Finally, given their tight behavioral connection, attempting to isolate a 'pure' form of uncertainty aversion risks reducing the construct to something less meaningful. In real-world reasoning, rigidity often stems from a deeper discomfort with uncertainty, making it both conceptually and practically relevant to the measurement of AOT.

Having said that, in Table 3, we showed an item measuring the avoidance of overconfidence.

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Option 'c' here requires a person not to be stubborn about his/her previous convictions, no matter how confident they were, but, in light of potential new evidence, to adjust his/her own levels of confidence and admit that they do not possess all the information and might be wrong. This is presumably hard to do, which is why this option will probably be chosen only by people who are serious about their AOT, that is, who do not have a problem in giving a fair treatment to evidence against their beliefs and revise their beliefs and confidence levels accordingly. Conversely, option 'a' indicates the lack of openness to counterevidence and changing beliefs/confidence accordingly. Option 'b' is again 'safe bet' in the middle of the 2 consisting of some elements of AOT.

In order to validate this initial AOT-SJT measure and choose items for further validation, we conducted a study on a community sample with the aim of investigating the psychometric properties of the measure and getting some initial evidence about its convergent validity. To do this, we recruited a sample of participants and asked them to solve our 20-item AOT-SJT together with a regular AOT scale and 2 subscales from the General Decision-Making Style questionnaire (GDMS; Scott and Bruce, 1995), measuring rational decision-making style that should positively correlate with AOT and intuitive decision-making style that should negatively correlate with AOT.

3.2. Method

3.2.1. Sample

A convenience sample of $N = 156$ respondents (57% women) participated in the study. Respondents were recruited by psychology students participating in the course on selection methods, and the only condition was that the participants were of legal age. Each student was instructed to recruit 2 participants and received course credits for the task. The participants' age ranged from 20 to 66 years ($M = 35.27$; $SD = 14.56$). As for the educational attainment, 0.7% of participants had completed primary education, 32.6% had completed secondary education, 25.8% had completed post-secondary non-tertiary education, 32.6% had completed tertiary education, and 8.3% held a doctoral or master's degree. No other information was collected.

3.2.2. Instruments

In addition to our 20-item-long AOT-SJT (6 items assessing the search quantity, 6 items assessing the search direction, and 8 items assessing the avoidance of overconfidence), we used:

AOT scale. We adopted an 11-item AOT scale from Baron et al. (2022) that measures the degree to which a person agrees with the AOT standards of thinking. Participants gave their level of agreement to statements such as 'People should take into consideration evidence that goes against conclusions they favor' and 'People should search actively for reasons why they might be wrong' on a 5-point scale, and the total score was calculated as the average of these ratings on all the items.

GDMS. We used two 5-item-long subscales of the GDMS (Scott and Bruce, 1995), one measuring Rational decision-making style (e.g., 'I make decisions in a logical and systematic way') and the other measuring Intuitive decision-making style (e.g., 'When I make decisions, I tend to rely on my intuition'), where participants rated their levels of agreement with each of the claims on a 5-point scale. The total scores on both subscales were again calculated as the average rating on all 5 items that comprise a subscale.

3.2.3. Procedure

The survey was constructed using the Guided track platform (https://www.guidedtrack.com/), where participants first gave their responses on our 20 AOT-SJT items, followed by the AOT scale and Rational and Intuitive decision-making subscales. At the beginning, participants were provided with a general explanation of the research purpose and an estimated completion time (15 minutes).

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Table 4. *Descriptive statistics and reliabilities of our measures.*

	M	*SD*	Observed range	Possible range	Omega total
AOT-SJT total	2.29	0.27	1.65–2.90	1.00–3.00	0.79
AOT-SJT search quant.	2.48	0.39	1.17–3.00	1.00–3.00	0.70
AOT-SJT search dir.	2.10	0.42	1.00–3.00	1.00–3.00	0.64
AOT-SJT overconf.	2.29	0.29	1.50–2.88	1.00–3.00	0.74
AOT scale	4.00	0.47	2.64–5.00	1.00–5.00	0.77
GDMS rational	4.25	0.59	1.80–5.00	1.00–5.00	0.85
GDMS intuitive	3.43	0.75	1.00–5.00	1.00–5.00	0.81

Table 5. *Pearson correlation coefficients between the variables used in the study.*

	AOT-SJT	AOT-SJT search quant.	AOT-SJT search dir.	AOT-SJT overconf.	AOT scale	GDMS rational	GDMS intuitive
AOT-SJT	–	/	/	/	.54	.45	−.34
AOT-SJT search quant.	.77***	–	.60	.38	.53	.35	−.31
AOT-SJT search dir.	.75***	.40***	–	.33	.23	.30	−.26
AOT-SJT overconf.	.67***	.27***	.23**	–	.52	.42	−.23
AOT scale	.42***	.39***	.16*	.39***	–	.49	−.42
GDMS rational	.37***	.27***	.22**	.33***	.40***	–	−.27
GDMS intuitive	−.27***	−.23**	−.19*	−.18*	−.33***	−.22***	–

Note: *$p < .05$, **$p < .01$, ***$p < .001$. The raw correlations are below the diagonal, whereas the disattenuated ones are above the diagonal.

3.3. Results

We first present the descriptive statistics of our measures, followed by a correlational analysis to see the convergent validity of the AOT-SJT measure, that is, how it relates to other variables we measured. Descriptive statistics are shown in Table 4, and the correlations are reported in Table 5.

Based on the descriptive analysis, we can observe that participants, on average, scored above the theoretical mean on AOT-SJT, meaning that they often choose an AOT-consistent response as the best one. Similarly, the raw mean score on the AOT scale was also high, indicating that participants, in general, agreed that AOT represents a standard for good thinking. The internal consistency of the AOT-SJT measure was satisfactory and similar to that of the AOT scale. The 3 subcomponents of AOT-SJT had somewhat lower internal consistency coefficients, which is not surprising, given that they are shorter.

Table 5 displays the correlation coefficients between the AOT-SJT instrument and other measures used in the study. AOT-SJT was moderately (or even strongly, if we consider the disattenuated correlation) correlated with AOT scale, meaning that those who agree with the principles of AOT have a greater tendency to behave in an AOT-consistent way in everyday situations. Furthermore, AOT-SJT correlations with rational and intuitive decision-making styles resembled those of the AOT scale—positive with rational and negative with intuitive decision-making styles. All these correlations testify to the validity of our new measure.

Given that SJT items are relatively lengthy and that solving 20 such items requires much time and mental effort, we wanted to try to shorten the measure before continuing with further validation in Study 2. To do this, we mainly relied on the corrected item-total correlations (correlations between the specific item and a full AOT-SJT score excluding that item) with the goal of excluding 2–3 worst-faring items per AOT-SJT subcomponent. In the end, we chose 13 items for Study 2, 4

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

from the search quantity and direction subcomponents, and 5 from the overconfidence avoidance subcomponent.[1]

4. Study 2

Through the second study, we aimed to build on the initial results from Study 1 and conduct additional analyses to validate the new AOT-SJT measure. Specifically, we wanted to additionally investigate the convergent and criterion validity of the AOT-SJT measure by correlating it with different constructs from its nomological network: the AOT scale, CRT, misinformation detection ability, overconfidence, and conspiracy mentality. Consistent with the literature review given in the introduction, we expected that the new AOT-SJT measure would fare similarly as the AOT scale in terms of correlations with other measures. Namely, we expected that it would correlate positively with the CRT score and misinformation detection ability, and negatively with overconfidence and conspiracy mentality. Of course, we expected that we would obtain a moderate positive correlation between the AOT-SJT and the AOT scale as we did in Study 1.

4.1. Method

4.1.1. Participants

A total of 379 Croatian participants took part in the study, with 55.7% identifying as female and 44.3% as male. The participants' ages ranged from 18 to 68 years ($M = 30.3$; $SD = 11.1$). Regarding educational attainment, 0.5% of participants had completed primary education, 36.9% had completed secondary education, 21.9% had completed post-secondary non-tertiary education, 33.7% had completed tertiary education, and 7.0% held a doctoral or master's degree. In terms of political orientation, the sample distribution skewed toward a more liberal orientation ($M = 2.51$ on a 1–5 scale; $SD = 0.89$), whereas in terms of religiosity, the sample exhibited a low level of religiosity ($M = 1.77$ on a 1–4 scale; $SD = 0.87$).

4.1.2. Instruments

AOT-SJT: A 13-item AOT-SJT measure was used, with 4 items measuring the direction of search, 4 items measuring the quantity of search, and 5 items tapping into the avoidance of overconfidence. Instructions and the scoring key remained the same. The responses range from 1 to 3, where 3 indicates the highest inclination toward AOT and 1 indicates the lowest inclination toward AOT. The overall score is expressed as the average value of the results on all items.

AOT scale: We used the same AOT scale version as in Study 1 (Baron et al., 2022), consisting of 11 items. The response scale and the total score calculation also remained the same.

CRT: Cognitive reflection was assessed using the items based on the original 3-item Cognitive Reflection Test (Frederick, 2005). Typical CRT task initially elicits an incorrect intuitive response, thereby measuring the inclination toward reflective thinking (Pennycook et al., 2020). An example item adapted for Croatian participants is 'A pencil and an eraser together cost 11 Croatian kunas. The pencil costs 10 kunas more than the eraser. How much does the eraser cost?' The intuitive response would be 1 kuna, but the correct answer is actually 0.5 kunas. The overall score is calculated as the average of the correct responses across all items.

Generic Conspiracist Beliefs scale (GCB): To measure conspiracy mentality, we used the short version of the GCB scale that contains a broad and representative range of conspiratorial beliefs (Brotherton et al., 2013). The scale consists of 15 items that describe beliefs in conspiracy theories well-known in scientific and popular literature. Examples of items include 'The government is involved in the murder of innocent citizens and/or well-known public figures and keeps this a secret' and 'The spread of certain viruses and/or diseases is the result of the deliberate, concealed efforts of some

[1]This had minimal effects on the reliability of AOT-SJT and its correlations with other measured variables, that is, the shortened and the full scale performed almost exactly the same.

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

organization'. The statements are assessed on a Likert-type scale, where 1 indicates 'Definitely not true' and 5 indicates 'Definitely true'. The items are formulated in a way that omits specific descriptions of particular organizations, events, or governmental bodies to remove context and ensure the test's generality and applicability across different cultures. The overall score is obtained by averaging the response values across all items.

Misinformation Susceptibility Test (MIST): We measured misinformation detection ability with the short version of the MIST developed by Maertens et al. (2021). The short version consists of 8 items, with 4 presenting true news headlines and 4 presenting false news headlines. Participants' task is to assess whether the presented headline is true or false. An example of a false headline item is 'Government officials manipulated stock prices to cover up scandals', whereas an example of a true headline item is 'Attitudes toward the EU are predominantly positive, both in Europe and beyond'. The test is scored based on participants' ability to distinguish between true and false headlines, and the overall score represents the proportion of correctly identified false and true headlines. A higher score indicates lower susceptibility to misinformation, that is, higher misinformation detection ability.

Overconfidence: A measure of overconfidence was derived in 2 steps. First, we followed the procedure reported in a study by Roozenbeek et al. (2022), where the authors, in addition to assessing the truthfulness of the headlines, asked participants to rate their level of confidence in their answers on MIST. The scale ranged from 1 to 7, where 1 represented 'Not at all confident in my response' and 7 represented 'Completely confident in my response', and we calculated the confidence index, which is a mean score on the confidence scale, and rescaled it to a range from 0 to 1 by dividing it by 7. In the second step, we used this information to obtain the overconfidence index by subtracting the participant's score on the MIST (mean accuracy) from the confidence index (previously calculated mean confidence scored rescaled to the 0–1 scale), resulting in the overconfidence index. This procedure is typical in studies on overconfidence, and this type of overconfidence is also known as overestimation (Moore and Healy, 2008).

At the end of the questionnaire, participants answered several questions about their sex, age, education level, religiosity, and political orientation. Religiosity was assessed on a scale from 1 to 4, where 1 means 'Not at all religious' and 4 means 'Extremely religious'. Political orientation was evaluated on a scale from 1 to 5, where 1 means 'Extremely left/liberal', 3 means 'Centrist', and 5 means 'Extremely right/conservative'.

4.1.3. Procedure

The study was conducted online using the Guided Track platform for creating interactive web surveys and applications. The sample was convenience-based and collected using the snowball sampling method through social media platforms Reddit and Facebook, as well as with the assistance of psychology students from the Faculty of Humanities and Social Sciences in Zagreb, who were tasked with recruiting 2 participants each, again as a part of course obligations. At the beginning, participants were provided with a general explanation of the research purpose, risks, and expected completion time, and asked to provide informed consent. Additionally, to motivate participants, we emphasized that they would receive feedback on their personal AOT score obtained from the AOT scale upon completion. The AOT-SJT was administered by presenting different response options to the instrument's items in a random order for different participants. Furthermore, at the end of the study, participants were provided with an explanation of the concept of active open-minded thinking and how they can foster it themselves.

4.2. Results

Before presenting the correlations between AOT-SJT and variables from its nomological network, we report on the descriptive statistics and reliabilities of our measures in Table 6.

Again, the average scores on AOT-SJT and AOT scale measures are above theoretical means, meaning that our participants agreed with AOT standards and tended to behave in AOT-consistent ways.

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Table 6. *Descriptive statistics and reliabilities of Study 2 measures.*

	M	*SD*	Observed range	Possible range	Omega total	$r(t_1-t_2)$
AOT-SJT	2.31	0.26	1.62–3.00	1.00–3.00	0.68	.64
AOT-SJT search quant.	2.58	0.36	1.50–3.00	1.00–3.00	0.55	.62
AOT-SJT search dir.	1.95	0.45	1.00–3.00	1.00–3.00	0.39	.58
AOT-SJT overconf.	2.38	0.31	1.40–3.00	1.00–3.00	0.52	.34
AOT scale	4.07	0.48	1.36–5.00	1.00–5.00	0.80	.78
CRT	0.59	0.38	0.00–1.00	0.00–1.00	0.86	/
Consp. Ment.	2.64	0.82	1.00–5.00	1.00–5.00	0.94	/
Misinfo.	0.69	0.19	0.00–1.00	0.00–1.00	0.70	/
Overconfidence	0.07	0.22	−0.56 to 1.00	−1.00 to 1.00	/	/

Note: M = mean; *SD* = standard deviation; $r\,(t_1 - t_2)$ = test–retest reliability; AOT-SJT = AOT situational judgment test; CRT = cognitive reflection test; Consp. Ment. = conspiracy mentality; Misinfo. = misinformation detection ability.

Table 7. *Pearson correlation coefficients between the variables used in the study.*

Instrument	AOT-SJT	AOT-SJT search quant.	AOT-SJT search dir.	AOT-SJT overconf.	AOT scale	CRT	Consp. Ment.	Misinfo	Overc.
AOT-SJT	–	/	/	/	.57	.24	−.28	.42	−.35
AOT-SJT search quant.	.72***	–	.73	.41	.47	.20	−.20	.36	−.31
AOT-SJT search dir.	.73***	.34**	–	.24	.34	.21	−.15	.27	−.19
AOT-SJT overconf.	.63***	.22**	.11*	–	.43	.16	−.20	.41	−.39
AOT scale	.42***	.31**	.19**	.28**	–	.35	−.43	.52	−.46
CRT	.18***	.14**	.12*	.11*	.29***	–	−.30	.23	−.27
Consp. Ment.	−.22***	−.14**	−.09	−.14**	−.37***	−.27***	–	−.60	.44
Misinfo	.29***	.22**	.14**	.25**	.39***	.18***	−.49***	–	−1
Overconfidence	−.29***	−.23**	−.12*	−.28**	−.41***	−.25***	.43***	−.85***	–

Note: *$p < .05$; **$p < .01$; ***$p < .001$. The raw correlations are below the diagonal, whereas the disattenuated ones are above the diagonal. AOT-SJT = AOT situational judgment test; CRT = cognitive reflection test; Consp. Ment. = conspiracy mentality; Misinfo. = misinformation detection ability.

This time, in addition to Omega total as an indicator of internal consistency, we also report the test–retest reliability for AOT-SJT and the AOT scale. We followed up with a subsample of our participants (N = 71) and asked them to solve these 2 measures again approximately 1 month after solving them the first time. We can see that the Omega total coefficient for the AOT-SJT measure is somewhat lower than the one obtained in Study 1, which was expected, given that the new measure is substantially shorter than the old one. Conversely, the Omega total coefficient for the AOT scale is roughly the same as in Study 1. The test–retest reliability was quite similar to the Omega total, both for AOT-SJT and the AOT scale. There were some differences between the 2 reliability indicators when looking at SJT subdimensions, where the SJT search direction had higher test–retest reliability compared with Omega total, whereas the opposite was true for SJT overconfidence. In general, the reliability of the AOT-SJT measure was in line with the average meta-analytically assessed internal consistencies (between .46 and .68; Catano et al., 2012; Kasten and Freund, 2016) or retest reliability of SJTs (.70; Harenbrock et al., 2023).

Table 7 displays the correlation coefficients between the AOT-SJT instrument and other measures used in the study to determine convergent, discriminant, and criterion validity.

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Convergent validity can be observed from the correlation between the AOT-SJT measure and the AOT scale, which is the strongest correlation between the measured variables. This moderate (strong when disattenuated) positive correlation basically replicates the results of Study 1, confirming that both AOT-SJT and the AOT scale tap into the construct of AOT. Additionally, AOT-SJT exhibited the same pattern of correlations with other variables as the AOT scale (although the correlations were somewhat weaker), positive with CRT and misinformation detection ability, and negative with conspiracy mentality and overconfidence. All of these correlations were significant and low-to-moderate in magnitude, confirming the convergent and criterion validity of the new AOT measure.

To further investigate the nature of the AOT-SJT criterion validity, we tested several regression models using structural equation modeling (SEM). Specifically, for each of the 3 criterion measures (conspiracy mentality, misinformation detection ability, and overconfidence), we compared a model where we regress the criteria on AOT-SJT and AOT scale scores (Model A) to a model where we regress the criteria to AOT-SJT and CRT scores (Model B). The idea of these analyses was to see whether AOT-SJT predicts the criteria because it taps into the disposition of AOT or only because it taps into cognitive abilities (as assessed by CRT). For example, if AOT-SJT does not exhibit incremental validity over the AOT scale score, but does over the CRT score, this would probably mean that AOT-SJT predicts these criteria due to the disposition to think and act in AOT-consistent ways.

We opted for the SEM approach, modeling and conducting the analyses on latent variables, because conducting the incremental validity analyses using ordinary regression suffers from serious drawbacks (e.g., inflated Type I errors) due to imperfect reliabilities of measures. Therefore, one of the possible solutions is to conduct regression analyses using SEM on latent variables that are free from measurement error (Westfall and Yarkoni, 2016). This way, if we observe the incremental validity of AOT-SJT over either the AOT scale or CRT, we can be more confident that this is not just a Type I error.

To conduct the SEM regression analyses, we specified models in which each latent variable was defined as a single factor, with all corresponding manifest variables loading onto it. We then regressed the latent outcome variable on the 2 latent predictor variables. For example, in the model examining whether the AOT scale and AOT-SJT predicted conspiracy mentality, we defined 3 latent variables: the AOT scale, measured by 11 manifest indicators; AOT-SJT, measured by 13 manifest indicators; and conspiracy mentality, measured by 15 manifest indicators. In all measurement models, the residuals of the manifest variables were specified as uncorrelated. Within the same model, we regressed the latent conspiracy mentality variable on the AOT scale and AOT-SJT latent variables to estimate the standardized beta coefficients for both predictors, as well as the total amount of variance explained in the outcome. The results of all SEM regression models are presented in Table 8.

Table 8 shows that β coefficients for AOT-SJT in Model B for all 3 criteria are larger than in Model A (and mostly significant, with the exception of Misinformation detection as an outcome for which AOT-SJT was almost significant). On the contrary, β coefficients for AOT-SJT in Model A are all nonsignificant. This generally means that cognitive abilities assessed with CRT are not the main reason for the predictiveness of AOT-SJT across the measured criteria, but that AOT-SJT predicts these criteria probably because it actually taps into AOT disposition. It is worth noting that not all model fit indices met conventional thresholds across all models, indicating some degree of model misfit. This primarily applied to the Comparative Fit Index (CFI), which, in some cases, fell below the commonly accepted cutoff of 0.90 (Hu and Bentler, 1999). Further inspection suggested that the misfit was largely attributable to correlations between the residuals of certain manifest variables that were not accounted for in the original model specification. Allowing some of these residuals to correlate typically improved model fit, raising the CFI above 0.90 in each case. However, because the standardized beta coefficients remained largely unchanged and the main conclusions unaffected, we report the more parsimonious models in which residuals were specified as uncorrelated.

However, the correlations given in Table 7 indicate that not all of the AOT-SJT subcomponents fare equally well in predicting other variables. Specifically, it seems that the search direction subcomponents fare worse than the other 2 in predicting 3 criteria variables: conspiracy mentality, misinformation detection ability, and overconfidence. Given that the direction of search (i.e., the myside bias avoidance)

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Table 8. *Results of SEM regression analyses investigating the incremental validity of AOT-SJT above the AOT scale score (Model A) and the CRT score (Model B) for the 3 criterion variables.*

	Criteria					
	Conspiracy mentality		Misinf. detection ability		Overconfidence	
	Model A β	Model B B	Model A β	Model B B	Model A β	Model B β
AOT-SJT	0.05	-0.25^{**}	0.17	0.54^{+}	−0.23	-0.39^{***}
AOT scale	-0.50^{***}		0.56*		−0.29*	
CRT		-0.27^{***}		0.17		-0.18^{**}
R^2	21.5%	16.9%	48.6%	36.7%	33.9%	21.9%
RMSEA	0.053	0.057	0.035	0.028	0.040	0.030
SRMR	0.059	0.057	0.053	0.051	0.054	0.048
CFI	0.805	0.838	0.785	0.860	0.808	0.908

Note: $^{***}p < .001$; $^{**}p < .01$; $^{*}p < .05$; $^{+}p = .067$. RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual; CFI = comparative fit index.

is the core aspect of AOT (i.e., Baron, 2024), we thought that it would be useful to try to improve its measurement. This is what we tried to do in Study 3 by including 3 additional items tapping into search direction.

5. Study 3

In Study 3, we investigated the convergent and criterion validity of the improved, larger AOT-SJT search direction subcomponent. This new version consists of 4 items from Study 1 and 3 new items. The new items come from our 'Good Boss' test, an SJT that measures 5 core leadership competencies ('Best of Both Worlds: Merging Traditional and Construct-Based Approaches to Develop the Good Boss Situational Judgment Test', manuscript under preparation) and is intended for the selection and development of managers. One of these competencies is decision-making, and the items in SJT measuring it are based on AOT but focus only on its search direction component. The logic is that the manager's propensity to behave in AOT-consistent ways in work-related decision-making situations should be a good indicator of his/her decision-making quality. Our initial validation of the 3 SJT items on a reasonably large sample of managers (n = 212) and their subordinates (n = 590) showed that the total score on the 3 items correlated with subordinates rating of workplace decision making effectiveness (r = .22; $p < .01$). Thus, these items are conceptually similar to the ones from current AOT-SJT, but with scenarios from the work realm, which is why we could just test them along with the existing ones.

5.1. Method

5.1.1. Participants

A convenience sample of N = 128 participants (Croatian citizens, 70% females) participated in our study. The mean age of the sample was M = 27.61 (SD = 11.08). Regarding the religiosity, the mean result was M = 1.90 (SD = 0.88) on a 4-point scale (1 = 'Not at all religious', 4 = 'Very religious'). Ideologically, our sample leaned left, scoring M = 2.38 (SD = 0.95) on a 5-point ideology scale (1 = 'Extremely left/liberal', 5 = 'Extremely right/conservative'). In terms of Education, 1.6% completed only elementary school, 49.2% finished high school, 27.3% had a bachelor's degree, 18.0% a master's degree, and 5.5% had postgraduate education.

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Table 9. *Descriptive statistics and reliabilities of Study 3 measures.*

	M	*SD*	Observed range	Possible range	Omega total
AOT-SJT dir.	2.17	0.34	1.29–3.00	1.00–3.00	0.67
AOT scale	4.03	0.48	2.36–5.00	1.00–5.00	0.76
CRT	0.54	0.41	0.00–1.00	0.00–1.00	0.91
NFC	3.48	0.90	1.00–5.00	1.00–5.00	0.85
FI	3.49	0.80	1.00–5.00	1.00–5.00	0.87
Consp. Ment.	2.93	0.84	1.00–5.00	1.00–5.00	0.77
Misinfo.	3.31	0.75	1.00–4.75	1.00–5.00	0.68

5.1.2. Instruments

AOT-SJT: We tested a version consisting of 7 items designed to tap into AOT-consistent tendency to search for the other-sided evidence and arguments. The logic of responses is similar to before, with the least AOT-consistent behavior being awarded one, and the most AOT-consistent behavior 3 points. The total score is calculated as the average of scores on all 7 items and can, thus, range from 1 to 3.

AOT scale: We used the same AOT questionnaire as in the previous 2 studies (Baron et al., 2022).

CRT: We used a 3-item CRT (Frederick, 2005), similar to the one we used in Study 2.

Rational-Experiential Inventory—Short (REI): We used a short-form REI (Norris et al., 1998) to capture 2 cognitive styles, NFC that emphasizes a conscious, analytical approach, and faith in intuition (FI) that emphasizes a pre-conscious, affective, holistic approach to thinking and making judgments (5 items each). The participants' task was to rate their levels of agreement with statements such as 'I would prefer complex to simple problems' (NFC) or 'My initial impressions of people are almost always right' (FI) on a 5-point scale (1 = 'Completely disagree', 5 = 'Completely agree'). The total score for both styles is calculated as the average of these ratings on all 5 items.

Generic Conspiracist Beliefs scale (GCB): To measure conspiracy mentality, we selected 4 out of 15 items from Brotherton et al.'s (2013) GCB scale (e.g., 'The spread of certain viruses and/or diseases is the result of the deliberate, concealed efforts of some organization'). The statements were assessed on a Likert-type scale, where 1 indicates 'Definitely not true' and 5 indicates 'Definitely true', and the overall score is obtained by averaging the response values across all items.

Misinformation Susceptibility Test (MIST): We measured misinformation detection ability again with the short version of the MIST developed by Maertens et al. (2021). This time, we used only 4 fake-news items and omitted real-news items.

5.1.3. Procedure

Once again, the questionnaire is administered online through the Guided Track platform. The participants were recruited with the help of psychology students who were given course credits. The AOT-SJT validation part was part of bigger data collection efforts for several research projects not related to the current one. Participants were informed at the beginning about all the tasks and questionnaires they would be solving and told that there are no particular risks related to the study and that they are free to opt out at any time. After providing us with informed consent, participants were able to continue with the questionnaire.

5.2. Results

We will again start by reporting the descriptive statistics of our measures, followed by the correlations among them. The descriptive statistics are shown in Table 9, and the correlations are reported in Table 10.

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Table 10. *Correlations among Study 3 variables.*

	AOT-SJT dir.	AOT scale	CRT	NFC	FI	Consp. Ment.	Misinfo.
AOT-SJT dir.	–	.56	.42	.33	−.12	−.18	.37
AOT scale	.40***	–	.46	.44	−.30	−.30	.58
CRT	.33***	.38***	–	.18	−.27	−.33	.36
NFC	.25***	.35***	.16	–	−.27	−.22	.41
FI	−.09	−.24**	−.24**	−.23*	–	.35	−.30
Consp. Ment.	−.13	−.23*	−.28**	−.18*	.29**	–	−.84
Misinfo.	.25**	.42***	.28**	.31***	−.23**	−.61***	–

Note: *$p < .05$; **$p < .01$; ***$p < .001$. The raw correlations are below the diagonal, whereas the disattenuated ones are above the diagonal.

Descriptive results for the AOT-SJT and AOT scale measures are quite similar to those obtained in previous studies. Again, mean scores on both scales were above the theoretical means, indicating that participants mostly agreed with AOT principles and tended to choose AOT-consistent responses in hypothetical scenarios. The reliabilities were again mostly good or acceptable, except for the somewhat lower internal consistency of the AOT-SJT measure that was expected, given the nature of the instrument.

Table 10 shows that the correlation between AOT-SJT and the AOT scale was again moderate and positive (strong after the disattenuation), which is consistent with the 2 previous studies and speaks in favor of the convergent validity of a new AOT measure. The correlations between AOT-SJT and other measures were similar in direction, if somewhat lower in magnitude, to those between the AOT scale and those measures. Compared with Study 2, we can see that the new and expanded AOT-SJT for search direction had stronger correlations in the expected direction with variables it was expected to correlate with (for the AOT scale: $r = .40$ vs. $r = .19$, for CRT: $r = .33$ vs. $r = .12$; and for conspiracy mentality: $r = -.20$ vs. $r = -.09$). Taken together, these results indicate that AOT-SJT for the search direction does manage to capture the tendency toward AOT and AOT-consistent behavior and that adding new items did improve the functioning of the measure.

Although we obtained encouraging results related to the validity of our new AOT-SJT measure in the first 3 studies, these studies were nevertheless conducted on the samples of Croatian participants. Thus, we decided to conduct the fourth and final study, whose aim was twofold. First, we wanted to examine the validity of AOT-SJT on the English-speaking population, which included the translation of items. Second, we wanted to test the full 16-item-long AOT-SJT measure, which includes all the items from Study 2 and new items tested in Study 3.

6. Study 4

6.1. Method

6.1.1. Participants

For our final study, we recruited a total of $N = 173$ U.S. participants using the Prolific platform. Our participants had a mean age of $M = 41.10$ years ($SD = 11.62$), were about equally split by gender (55% females), and mostly had a college degree (48%), followed by high school (31%) and Master's or PhD (21%). Regarding ideology, our participants scored $M = 4.81$ ($SD = 1.76$) on a scale where 1 = extremely left/liberal and 10 = extremely right/conservative, meaning that, on average, they were quite close to being in the center, with a respectable number of participants on both sides of the ideological spectrum.

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Table 11. *Descriptive statistics and reliabilities of the measures used in Study 4.*

	M	*SD*	Observed range	Possible range	Omega total
AOT-SJT	2.31	0.21	1.75–2.88	1.00–3.00	0.68
AOT-SJT quant.	2.51	0.35	1.25–3.00	1.00–3.00	0.53
AOT-SJT dir.	2.13	0.30	1.29–2.86	1.00–3.00	0.61
AOT-SJT overconf.	2.40	0.28	1.60–3.00	1.00–3.00	0.40
AOT scale	4.01	0.49	2.82–5.00	1.00–5.00	0.88
CRT	0.44	0.36	0.00–1.00	0.00–1.00	0.81
GDMS rational	4.24	0.57	1.60–5.00	1.00–5.00	0.91
GDMS intuitive	3.54	0.71	1.60–5.00	1.00–5.0	0.89

6.1.2. Instruments

AOT-SJT: Our final version of AOT-SJT consisted of 16 items in total, 7 items measuring the direction of search (the same items used in Study 3), 4 items measuring the quantity of search, and 5 items tapping into the avoidance of overconfidence (the latter 2 dimensions were captured with items used in Study 2). The full instrument translated into English is available in the Appendix.

AOT scale: We used the same AOT scale version as in previous studies (Baron et al., 2022), consisting of 11 items. The response scale and the total score calculation also remained the same.

CRT: We used the original, 3-item CRT version (Frederick, 2005), very similar to the one we used in Study 2, only this time in English instead of Croatian.

GDMS: Finally, just as in Study 1, we used two 5-item-long subscales of the GDMS (Scott and Bruce, 1995), one measuring Rational decision-making style and the other measuring Intuitive decision-making style.

6.2. Procedure

Prior to running the survey, we translated our AOT-SJT measure from Croatian to English. The translation was done by 2 authors (N.E. and Z.G.) who are fluent in English, and the final version is the one that both authors agreed on. The full questionnaire was administered online through the Guided Track platform, and participants were recruited via the Prolific platform and paid for their participation. Participants were informed at the beginning about all the tasks and questionnaires they would be solving, about the potential risks, and told they were free to opt out at any time. After providing informed consent, participants were able to continue with the questionnaire.

6.3. Results

We calculated the descriptive statistics and the reliabilities of our variables, as well as the correlations between them. We are showing the descriptive statistics and reliabilities of our measures in Table 11, whereas the correlations between them are shown in Table 12.

Although our Study 4 results are mostly in line with expectations (i.e., AOT-SJT and its subscales correlate positively with AOT scale and, to some extent, with CRT, and negatively with intuitive decision-making), what catches the eye is that these correlations are markedly lower than was the case in the previous studies. For example, the AOT-SJT correlation with the AOT scale of $r = .27$ (disattenuated $r = .35$) is smaller than the same correlation obtained in the previous 3 studies ($r = .41$ in Study 1, $r = .42$ in Study 2, and $r = .40$ in Study 3). There are 2 possibilities that come to mind here. First, it is possible that our SJT items function somewhat differently in the U.S. population, that

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Table 12. *Correlations between the Study 4 variables.*

	AOT-SJT	SJT search quant.	SJT search dir.	SJT overconf.	AOT scale	CRT	GDMS rational	GDMS intuitive
AOT-SJT	–	/	/	/	.35	.20	−.09	−.23
AOT-SJT search quant.	.57***	–	.35	.02	.22	.17	−.01	−.20
AOT-SJT search dir.	.81***	.20**	–	.47	.25	.17	−.17	−.19
AOT-SJT overconf.	.58***	.01	.23**	–	.35	.07	.10	−.08
AOT scale	.27***	.15*	.18*	.21**	–	.34	.32	−.37
CRT	.15	.11	.12	.04	.29***	–	.08	−.29
GDMS rational	−.07	−.01	−.13	.06	.29***	.07	–	−.12
GDMS intuitive	−.18*	−.14	−.14	−.05	−.33***	−.25**	−.11	–

Note: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. The raw correlations are below the diagonal, whereas the disattenuated ones are above the diagonal.

is, that our SJT items are not that strong an indicator of open-minded thinking/behavior, at least not for this particular U.S. sample. For instance, notable cultural differences have been observed between Croatia and the United States (Rajh et al., 2016; Tavakoli et al., 2003). Croatian citizens tend to score higher than their U.S. counterparts on uncertainty avoidance—that is, the degree to which individuals feel threatened by ambiguous or unknown situations. In contrast, they score lower on masculinity (the extent to which cultural values emphasize traits traditionally associated with masculinity, such as competitiveness and dominance) and individualism (the extent to which individual interests are prioritized over group interests). These cultural differences may influence how AOT is expressed. For example, lower levels of masculinity may foster a stronger preference for consensus and conflict avoidance, potentially making individuals more receptive to others' viewpoints. In such cultural contexts, agreement with AOT principles might more readily translate into inclusive and open-minded behaviors. Conversely, in cultures with higher masculinity, dominant norms that value assertiveness and competition could inhibit the expression of AOT-consistent behavior, even among individuals who endorse AOT values in principle. These propositions remain speculative, however, as identifying the specific mechanisms by which cultural values shape AOT expression lies beyond the scope of this article.

The second possible reason for the lower correlation between the AOT-SJT and the AOT scale is that Prolific participants were not that attentive and careful when reading and solving SJT items. Although we had 2 attention check questions that were passed by all but 2 participants, the participants might have been trained to quickly read and comprehend relatively short items, unlike the longer SJT items, for which they might not have had patience or time to read and solve carefully. Notwithstanding these possibilities, we believe that Study 4 results align with the results of other studies, providing initial evidence that AOT-SJT might serve as a promising new measure of AOT.

In addition to the analyses we presented within the manuscript, we also provide the correlations of each AOT-SJT item with the AOT scale and CRT on a pooled sample of all 4 studies in the Appendix, as we believe that this might be helpful for researchers who might wish to use only some of the items from our AOT-SJT in the future. Additionally, based on the pooled sample, we included Figure A1 in the Appendix, which displays AOT scale scores for participants based on their selected response option ('a', 'b', or 'c') for each SJT item. The figure shows a consistent pattern: participants who selected option 'c' (the response we considered most aligned with AOT) scored highest on the AOT scale, followed by those who chose option 'b', and then those who chose option 'a' (the response we considered least aligned with AOT). This pattern not only supports the conceptual rationale behind our ordering of the response options but also provides empirical validation for it.

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

7. Discussion

If AOT represents a standard for rationality (Baron et al., 2022), then to be rational, one must behave in AOT-consistent ways. For example, when a person concludes that (s)he needs to collect some more information before coming to a conclusion or making a judgment or decision, (s)he must perform such a search in a fair manner, by searching not only information/evidence that corroborates their preferred position, but also those that oppose it. Furthermore, if one is to act rationally, then one must, after encountering counterevidence, seriously consider it and not downplay it. Of course, one must then fairly incorporate opposite evidence in his/her thinking by either changing the conclusion or decision, or at least adjusting the confidence in such conclusions/decisions. These few sentences probably sound trivial because they follow straight from the AOT theory, yet, at the same time, they accentuate the problems with the current ways we assess AOT, namely with AOT scales.

The AOT scale asks participants whether they agree in principle with the statements describing how good thinking should look like. By looking at AOT scale scores, we learn about the degree to which an individual agrees that AOT is a proper standard for the quality of thinking. However, there are several things that we do not learn. We do not learn (a) whether an individual actually applies these principles to his/her own thinking; (b) whether an individual recognizes the need to reason in AOT way in a specific situation; and (c) whether an individual knows how to apply these principles in a range of situations (s)he might be faced with. We write this not to downplay the AOT scale that is almost exclusively used to assess AOT today, as it clearly performs well and is able to predict a wide range of phenomena, as described in the introduction. We stress this to exemplify the areas we thought a new measure could address. This was a main driver behind our efforts to construct a new AOT-SJT measure—to try to address the issues with the current scale by constructing an SJT measure that bypasses these issues and to provide an alternative AOT measure that would be relatively easy to implement and use.

In short, we believe that our results across the 4 studies consistently show that AOT-SJT is a promising new way of capturing AOT. Notably, in all 4 studies, it exhibited a moderate-to-high positive correlation with the AOT scale. Some might view this as insufficient, as generally higher effects are expected if one is to claim that 2 measures capture the same thing. However, there are multiple reasons why a stronger correlation would be hard to achieve. The most obvious is the method factor—these 2 measures capture certain traits using different methods. For example, Olaru et al. (2019) validated an SJT for measuring dependability as a facet of conscientiousness, compared it to standard measures of conscientiousness, and observed that method-related variance accounted for as much as 12.25% of the total variance. In addition to the method factor, probably a more important reason for imperfect correlation between the measures is that, as we mentioned before, the AOT-SJT and the AOT scale actually do not capture exactly the same construct or identical parts of the same construct. As Bledow and Frese (2009) explained, a situational heterogeneity of items introduces specific variance into each SJT item. Therefore, it is not only the proclivity to AOT that defines how one will respond to a specific item, but also one's beliefs about whether the situation requires AOT-consistent behavior, one's previous experiences in similar situations and whether one has a procedural knowledge of how to behave in an AOT-consistent way in a specific situation, an important aspect of SJTs (Lievens, 2017; Lievens and Motowidlo, 2016).

This aligns with ongoing discussions about the often modest correlations between self-report and behavioral measures of psychological constructs. Dang et al. (2020) highlight that such discrepancies stem in part from differences in response processes: behavioral measures typically capture responses on specific, structured situations and are often scored objectively (e.g., for accuracy), whereas self-reports reflect individuals' subjective judgments about their behaviors or beliefs across diverse, unstructured situations. Although the SJT is not a purely behavioral measure—it requires 'only' reasoning about and inferring the most appropriate actions rather than performing them—it shares key features with behavioral tasks, such as standardized stimuli and accuracy-based scoring. These differences also likely contribute to the reduced correlation between the AOT scale and AOT-SJT scores.

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Given all these sources of variability, together with the fact that we purposely sampled items from a wide range of domains (e.g., work life, family life, housing, health, and hobbies), we think that the moderate-to-high correlations between the general AOT scale and situation-specific AOT-SJT are quite encouraging. In a way, they testify that people who say that AOT is a proper way of thinking actually show intentions to behave in an AOT-consistent way.

Other correlations between AOT-SJT and conceptually related variables from its nomological network generally confirm that AOT-SJT is a valid AOT measure. The direction, and to some extent the magnitude, of these correlations resemble those of the AOT scale, showing that the new AOT measure behaves similarly to the established one. Furthermore, in the biggest study, Study 2, we provided evidence that its criterion validity is not due to cognitive ability (as assessed with CRT), but mostly due to capturing the tendency to think and behave in an AOT way.

Before discussing the possible uses and benefits of the new measure, we will briefly touch on its differences from existing AOT scales and comment on some issues with the new measure. We already noted the general differences between the AOT-SJT and the AOT scale, but will reiterate the most important ones. Unlike the AOT scale, following Bledow and Frese (2009), we see AOT-SJT as a formative measure. This means that the construct is defined by its manifest indicators (instead of existing independently and influencing the responses on manifest variables, i.e., being reflected in them), which further means that it will be defined and measured as good as manifest variables are good themselves. In this situation, this means that our AOT-SJT measure will be a good measure of AOT if there exists a broad and representative sample of situations in which AOT-consistent behavioral intentions can be expressed. This also presents an area for future improvement of the AOT-SJT by adding new or changing/removing old situations.

Another main difference between the AOT scale and AOT-SJT is that SJT should largely tap into procedural knowledge (Lievens, 2017; Lievens and Motowidlo, 2016), that is, into testing whether a person knows how to behave in an AOT way in a specific situation, while the AOT scale measures generalized tendency to agree with AOT thinking. To be clear, we believe that this tendency actually translates into thinking and behaving in an AOT way; otherwise, the AOT scale would probably not be correlated with such a wide range of phenomena. The fact that AOT-SJT captures parts of the construct that the AOT scale probably misses, but is also restrained by the breadth and quality of situations, while the AOT scale is wider, capturing generalized tendency, but probably not completely translating into actual behavior, means that these 2 measures could be seen as complementary.

At this point, it is appropriate to comment on some of the issues with the new AOT-SJT and the lessons we learned during its development. First, one of the more serious issues with the AOT-SJT is its modest reliability. There is a clear trade-off between the situational breadth and richness of the measure and its psychometric properties. In the case of SJTs, this is not surprising. Each item essentially has its own specific latent variable representing a person's true behavioral preference in a particular simulated situation (MacKenzie et al., 2005). This situation-specific variance naturally limits the reliability of such measures, a well-documented finding in the SJT literature (with internal consistency coefficients ranging from .46 to .68 in meta-analyses by Catano et al., 2012, and Kasten and Freund, 2016). This lower reliability may weaken the observed effect sizes between AOT-SJT and other variables, although there are possible ways to address this issue.

The most obvious approach is to increase the number of items, but this quickly becomes impractical and time-consuming for participants. Another possibility is to use a Likert-type scale for each response option, allowing participants to indicate the likelihood of acting in each possible way. This approach could improve internal consistency (e.g., Peus et al., 2013). A further option would be to reduce situation-specific variance by keeping the wording of items as similar as possible or by drawing all situations from a single domain. However, given our earlier discussion on the importance of situational diversity, we are not convinced that these approaches would be either particularly fruitful or preferable. Nonetheless, future research could explore these possibilities.

Second, we learned that it is challenging to develop situations that are perceived uniformly across participants, and this subjectivity in perception can undermine validity. In other words, as noted

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

earlier, how someone interprets a situation affects whether they respond in the way we consider AOT-consistent, regardless of their actual disposition toward AOT. There are at least 2 situational features that contribute to this challenge: (a) specificity of the situation—in an attempt to create realistic scenarios, we often include many details, which can cue idiosyncratic interpretations and responses, thereby influencing answers beyond participants' true AOT levels; and (b) perceived importance or seriousness—not all individuals evaluate situations according to the same criteria. Some may consider a situation important and worth deeper thinking, whereas others may not see it as worthy of their time or cognitive effort, leading them not to engage in AOT. This variability can erode the validity of the measure, as responses are influenced by factors other than a genuine inclination toward AOT. In this study, we attempted to control for this by selecting only situations involving important decisions. However, importance is inherently subjective. In future studies, it might be useful to explicitly control for this factor by pre-testing situations and selecting those that a majority of people agree are either important and worth thoughtful consideration or unimportant and not worth the effort.

A third issue is the potential for cultural or normative differences in trait expression across countries, which can hinder the generalizability and cross-cultural validity of the AOT-SJT. Motivated by the lower observed validity of the AOT-SJT in the U.S. sample compared to Croatian samples, we speculated that cultural differences between the 2 countries might influence how the trait is expressed. In other words, cultural and social norms can vary significantly across countries, leading individuals with similar levels of AOT disposition to express the trait in different ways. This variability can, in turn, decrease or even diminish the measure's validity. This is an important avenue for future research and warrants more careful exploration.

Finally, we see several uses for the new AOT-SJT measure. For example, it can be used either as an alternative or as a complementary AOT measure, as each addresses the shortcomings of the other. For example, we previously used a short 3-item AOT-SJT from 'The Good Boss' test (reported in Study 3) alongside the AOT scale and observed that it exhibited similar correlations as the AOT scale with different measures of cognitive biases, personality traits, and decision-making quality as assessed by peers (Erceg et al., 2022). Furthermore, AOT-SJT could be used as a better indicator of actual, rather than just proclaimed, AOT, and, finally, it could be used as a dependent variable to see whether one's AOT-related behavioral intentions change over time or after some specific intervention. As an example for the latter use, we applied a 13-item long AOT-SJT (from Study 2) on the same sample of participants at 3 different time points when testing the effectiveness of 2 educational interventions to teach AOT and observed that the AOT-SJT scores increased after the interventions (Erceg et al., 2024). Therefore, we hope that our introduction and initial validation of AOT-SJT will encourage other researchers to use it and improve it in the future.

8. Conclusion

Across 4 studies, we developed and validated a new AOT-SJT measure. We showed that it possesses promising psychometric qualities and that it exhibits convergent and criterion validity similar to that of an established AOT scale. The new AOT measure differs in substantial and important ways from the existing scale, but we believe that it complements it well and provides a viable alternative for assessing AOT-consistent thinking and behavioral tendency.

Data availability statement. Data for all 4 studies are available at https://osf.io/8ba2s/?view_only=386bc2d8ad914d77aa75ff8ce69e866a.

Funding statement. Research in this article was partially funded through the Croatian Science Foundation (Grant No. IP-2022-10-3356).

Competing interest. The authors declare no competing interests.

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

References

Athanasiou, K., & Papadopoulou, P. (2012). Conceptual ecology of the evolution acceptance among Greek education students: Knowledge, religious practices and social influences. *International Journal of Science Education*, *34*(6), 903–924. https://doi.org/10.1080/09500693.2011.586072

Baron, J. (1985). *Rationality and intelligence*. Cambridge University Press. https://doi.org/10.1017/CBO9780511571275

Baron, J. (1991). Beliefs about thinking. U knjizi J. F. Voss, D. N. Perkins i J. W. Segal (Eds.), *Informal reasoning and education*. Lawrence Erlbaum Associates, Inc.

Baron, J. (2018). Social norms for citizenship. *Social Research: An International Quarterly*, *85*(1), 229–253. https://doi.org/10.1353/sor.2018.0011

Baron J. (2019). Actively open-minded thinking in politics. *Cognition*, *188*, 8–18. https://doi.org/10.1016/j.cognition.2018.10.004

Baron, J. (2023). *Thinking and deciding*. Cambridge University Press.

Baron, J. (2024). Two components of individual differences in actively open-minded thinking standards: Myside bias and uncertainty aversion. *Thinking & Reasoning*, *30*(4), 648–673.

Baron, J., Isler, O., & Yilmaz, O. (2022). *Actively open-minded thinking and the political effects of its absence*. PsyArXiv. https://doi.org/10.31234/osf.io/g5jhp

Bledow, R., & Frese, M. (2009). A situational judgment test of personal initiative and its relationship to performance. *Personnel Psychology*, *62*(2), 229–258.

Brotherton, R., French, C. C., & Pickering, A. D. (2013). Measuring belief in conspiracy theories: The Generic Conspiracist Beliefs scale. *Frontiers in Psychology*, *4*, 279. https://doi.org/10.3389/fpsyg.2013.00279

Catano, V. M., Brochu, A., & Lamerson, C. D. (2012). Assessing the reliability of situational judgment tests used in high-stakes situations. *International Journal of Selection and Assessment*, *20*(3), 333–346.

Dang, J., King, K. M., & Inzlicht, M. (2020). Why are self-report and behavioral measures weakly correlated? *Trends in Cognitive Sciences*, *24*(4), 267–269.

Ellis, G. (2018). So, what are cognitive biases? Ellis, G. (Ed.) *Cognitive biases in visualizations*. Springer, 1–10. https://doi.org/10.1007/978-3-319-95831-6_1

Erceg, N., Andrić, L., Bosilj, L., Britvić, F. D., Čeko, A., Dedić, M., . . . Galic, Z. (2024). Teaching actively open-minded thinking online: Encouraging effects of a serious computer game and an online module. *Journal of Cognitive Psychology*, *36*(8), 938–953. https://doi.org/10.31234/osf.io/hwzk7

Erceg, N., Galić, Z., & Bubić, A. (2022). Normative responding on cognitive bias tasks: Some evidence for a weak rationality factor that is mostly explained by numeracy and actively open-minded thinking. *Intelligence*, *90*, 101619.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Haran, U., Ritov, I., & Mellers, B. A. (2013). The role of actively open-minded thinking in information acquisition, accuracy, and calibration. *Judgment and Decision Making*, *8*, 188–201.

Harenbrock, J., Forthmann, B., & Holling, H. (2023). Retest reliability of situational judgment tests: A meta-analysis . *Journal of Personnel Psychology*, *22*(4), 169–184. https://doi.org/10.1027/1866-5888/a000323

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Janssen, E. M., Verkoeijen, P. P., Heijltjes, A. E., Mainhard, T., van Peppen, L. M., & van Gog, T. (2020). Psychometric properties of the actively open-minded thinking scale. *Thinking Skills and Creativity*, *36*(6), 100659. https://doi.org/10.1016/j.tsc.2020.100659

Kasten, N., & Freund, P. A. (2016). A meta-analytical multilevel reliability generalization of situational judgment tests (SJTs). *European Journal of Psychological Assessment*, *32*(3), 230–240. https://doi.org/10.1027/1015-5759/a000250

Lievens, F. (2017). Construct-driven SJTs: Toward an agenda for future research. *International Journal of Testing*, *17*(3), 269–276. https://doi.org/10.1080/15305058.2017.1309857

Lievens, F., & De Soete, B. (2015). Situational judgment tests. In *International encyclopedia of the social & behavioral sciences*. *22*, 13–19, Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.25092-7

Lievens, F., & Motowidlo, S. J. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology*, *9*(1), 3–22.

MacKenzie SB, Podsakoff PM, & Jarvis CB. (2005). The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology*, *90*(4), 710–730.

MacLaren, V. V., Fugelsang, J. A., Harrigan, K. A., i Dixon, M. J. (2012). Effects of impulsivity, reinforcement sensitivity, and cognitive style on pathological gambling symptoms among frequent slot machine players. *Personality and Individual Differences*, *52*(3), 390–394. https://doi.org/10.1016/j.paid.2011.10.044

Maertens, R., Götz, F. M., Schneider, C. R., Roozenbeek, J., Kerr, J. R., Stieger, S., McClanahan III, W. P., Drabot, K., & Linden, S. (2021). *The Misinformation Susceptibility Test (MIST): A psychometrically validated measure of news veracity discernment.* PsyArXiv. https://doi.org/10.31234/osf.io/gk68h

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, *60*(1), 63–91. https://doi.org/10.1111/j.1744-6570.2007.00065.x

Milosavljevic, M. M., Koch, C., & Rangel, A. (2011). Consumers can make decisions in as little as a third of a second. *Judgment and Decision Making*, *6*(6), 520–530.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502.

Norris, P., Pacini, R., & Epstein, S. (1998). The Rational-Experiential Inventory, Short Form. Unpublished inventory. University of Massachusetts at Amherst.

Nutt, P. C. (1999). Surprising but true: Half the decisions in organizations fail. *Academy of Management Perspectives*, *13*(4), 75–90. https://doi.org/10.5465/ame.1999.2570556

Olaru, G., Burrus, J., MacCann, C., Zaromb, F. M., Wilhelm, O., & Roberts, R. D. (2019). Situational judgment tests as a method for measuring personality: Development and validity evidence for a test of dependability. *PLoS One*, *14*(2), e0211884. https://doi.org/10.1371/journal.pone.0211884

Oostrom, J. K., de Vries, R. E., & De Wit, M. (2019). Development and validation of a HEXACO situational judgment test. *Human Performance*, *32*(1), 1–29. https://doi.org/10.1080/08959285.2018.1539856

Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2020). On the belief that beliefs should change according to evidence: Implications for conspiratorial, moral, paranormal, political, religious, and science beliefs. *Judgment & Decision Making*, *15*(4), 476–498. https://doi.org/10.31234/osf.io/a7k96

Peus, C., Braun, S., & Frey, D. (2013). Situation-based measurement of the full range of leadership model—Development and validation of a situational judgment test. *The Leadership Quarterly*, *24*(5), 777–795. https://doi.org/10.1016/j.leaqua.2013.07.006

Rajh, E., Budak, J., & Anić, I. D. (2016). Hofstede's culture value survey in Croatia: Examining regional differences. *Društvena Istraživanja*, *25*(3), 309–327.

Roozenbeek, J., Maertens, R., Herzog, S. M., Geers, M., Kurvers, R. H., Sultan, M., & van der Linden, S. (2022). Susceptibility to misinformation is consistent across question framings and response modes and better explained by myside bias and partisanship than analytical thinking. *Judgment and Decision Making*, *17*(3), 547–573.

Scott, S. G., & Bruce, R. A. (1995). Decision-making style: The development and assessment of a new measure. *Educational and Psychological Measurement*, *55*(5), 818–831.

Sibony, O. (2020). *You're about to make a terrible mistake!: How biases distort decision-making and what you can do to fight them*. Swift Press.

Stankov, L., & Lee, J. (2014). Overconfidence Across World Regions. *Journal of Cross-Cultural Psychology*, *45*(5), 821–837. https://doi.org/10.1177/0022022114527345

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, *89*(2), 342–357. https://doi.org/10.1037/0022-0663.89.2.342

Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning*, *13*(3), 225–247. https://doi.org/10.1080/13546780600780796

Svedholm-Häkkinen, A. M., & Lindeman, M. (2018). Actively open-minded thinking: Development of a shortened scale and disentangling attitudes towards knowledge and people. *Thinking & Reasoning*, *24*(1), 21–40. https://doi.org/10.1080/13546783.2017.1378723

Swami, V., Coles, R., Stieger, S., Pietschnig, J., Furnham, A., Rehim, S., & Voracek, M. (2011). Conspiracist ideation in Britain and Austria: Evidence of a monological belief system and associations between individual psychological differences and real-world and fictitious conspiracy theories. *British Journal of Psychology*, *102*(3), 443–463. https://doi.org/10.1111/j.2044-8295.2010.02004.x

Swami, V., Voracek, M., Stieger, S., Tran, U. S., & Furnham, A. (2014). Analytic thinking reduces belief in conspiracy theories. *Cognition*, *133*(3), 572–585. https://doi.org/10.1016/j.cognition.2014.08.006

Tavakoli, A. A., Keenan, J. P., & Cranjak-Karanovic, B. (2003). Culture and whistleblowing an empirical study of Croatian and United States managers utilizing Hofstede's cultural dimensions. *Journal of Business Ethics*, *43*, 49–64.

Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLoS One*, *11*(3), e0152719.

Zajenkowski, M., Górniak, J., Wojnarowski, K., Sobol, M., & Jonason, P. K. (2022). I need some answers, now!: Present time perspective is associated with holding conspiracy beliefs. *Personality and Individual Differences*, *196*, 111723. https://doi.org/10.1016/j.paid.2022.111723

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Appendix

Table A1. *All items from the final AOT-SJT (one that is tested in Study 4).*

No.	Item
1.	You work for a newly founded IT company, and this year, as part of promoting socially responsible business, the firm wants to donate a large sum of money to a charitable organization. Your work team has the task of deciding which charitable organization will receive these funds. It is very important to choose a verified and high-quality organization, with which it would be possible to establish an effective and sustainable partnership. The image of your company as a desirable and responsible employer, but also your own reputation, depend on this decision. You need to choose an organization within the next week, and this organization must operate within the area of your city. What should you do in this situation to make the best possible decision? 1. You will suggest to your team to donate to a well-known and popular organization that you saw on TV and which you have heard many other companies donate to. You believe that this is the best way to make a good decision within the given deadline. 2. You will consider a recommendation from a close friend you trust, whose wife runs a charity. After you get more details about the organization from him, you will propose it to your team. 3. After searching the Internet for numerous charities and writing down a few options that particularly impressed you, you will contact each of these organizations to personally verify the accuracy of the information and ask your team to help you make a final decision.
2.	Your 2-and-a-half-year-old child currently goes to a daycare center in the neighborhood where you live. In that same building, there is also a kindergarten. A friend whose child is enrolled in that kindergarten tells you that the kindergarten is excellent and that there is even the possibility of learning a foreign language. You are considering which kindergarten you should enroll your child in when he turns 3. What should you do in this situation to make the best possible decision? 1. Due to the proximity and extremely positive experiences of your friend, you will enroll your child in the kindergarten that is in the same building as the current daycare center. 2. You decide to talk to a few other parents whose children go to that kindergarten in order to hear their impressions and experiences before making a decision about where to enroll your child. 3. You have decided to take a week to research which kindergartens are located within a few kilometers of your place of residence. You will focus on the number of educators, available professionals within the facility, type of program (e.g., Montessori and Waldorf), proximity, and reviews.
3.	You have been working in your field for 3 years in an organization in which you were employed right after college. You are satisfied with your current job and relationships with colleagues, but occasionally encounter misunderstandings with your supervisor. You have received a job offer from another organization, and after successfully going through the selection process, you have been offered the same position with a 10% higher salary than your current one. They have asked for your response within 4 days. What should you do in this situation to make the best possible decision?

(Continued)

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Table A1. *(Continued).*

No.	Item
	1. Considering it's a good opportunity and offers a higher salary, you will accept the job immediately. A quick response will demonstrate motivation and decisiveness. If you don't respond promptly, the company might hire someone else. 2. You will have a conversation with an acquaintance who works at the organization that offered you the job. You will ask about their experiences and seek advice. After this discussion, you will make a decision as soon as possible. 3. You will express your gratitude about the offer and say that you will need to think about it. Over the next few days, you will thoroughly research the organization that offered you a job and try to gather information from their employees. You will consult your colleagues and consider the pros and cons of your current job, as well as the offered job before making a decision.
4.	Lately you've been suffering from insomnia and anxiety, you're constantly in a bad mood, and you don't know how to help yourself. Your neighbor recommended that you try a well-known alternative medicine that allegedly works well for cases of psychological distress and your neighbor swears that it will solve all your problems. Despite the fact that it is quite expensive, you know that you need to help yourself as soon as possible because the symptoms are becoming unbearable. What should you do in this situation to make the best possible decision? 1. You trust your neighbor's recommendation because you have known her for a very long time, and she herself tried the medicine she recommends, so you decide to buy it at the first opportunity. 2. Many people struggle with these kinds of problems, and you have heard positive experiences with similar treatments from some acquaintances. Therefore, you decide to call a few acquaintances to inquire about their experiences with this medicine because the experiences of others are a very good indicator of the effectiveness of the drug. 3. You search the Internet to find expert reviews and articles on the effects of this drug and its scientific basis. You will make a decision only after you are convinced that this medicine can actually help you.
5.	You lead the Human Resources Department in an organization, and you have heard that other companies are implementing flexible working hours for their employees. This sounds like a good and interesting idea to you because you believe it will ease the workload and enhance efficiency. You are considering proposing this idea to the CEO, but before doing so, you want to gather additional information. What should you do in this situation to make the best possible decision? 1. You will talk with colleagues from the HR departments of organizations that have had positive experiences after implementing such a working arrangement. You will ask them to explain the benefits of flexible working hours and how your organization can benefit from its introduction. 2. You will seek advice from an acquaintance whose organization has recently introduced flexible working hours. You will ask them about their experiences and the results of this approach. 3. You will try to identify organizations that haven't had positive experiences after implementing this kind of working arrangement. In discussions with colleagues from these organizations, you will pay special attention to potential drawbacks of flexible working hours.

(Continued)

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Table A1. *(Continued).*

No.	Item
6.	You dislike your nose and you've been thinking your whole life about reducing its size. During elementary and high school, you were often teased and called 'Long Nose', which made you always keep to the sidelines. Now that you're an adult, you have the financial means to finally undergo plastic surgery that has been on your mind for a long time. Although the benefits of the surgery personally appeal to you and outweigh any drawbacks you can think of, before making a decision, you want to gather additional information. What should you do in this situation to make the best possible decision? 1. In social media groups for people who have undergone a similar procedure, you will seek opinions and experiences from individuals who think that the procedure significantly improved their quality of life. You want to learn more about the advantages and positive effects of the surgery. 2. You will search for more information about the procedure on the Internet. Also, during consultations with a surgeon from a clinic, you will ask them to provide you with a detailed overview of the surgery process and recovery. 3. By searching the Internet and social media, you will attempt to find negative patient experiences and key arguments against your decision to undergo the surgery.
7.	The school psychologist has invited you to talk about your child. She tells you that it has been noticed that your child shows signs of giftedness. Therefore, the psychologist makes a proposal for the child to skip a grade and join a long-term program for gifted children, which includes a range of activities and a school program that is a third more extensive than the standard one. This information made you happy and the proposed program seems like a good way to encourage and develop your child's abilities. However, before making a decision you want to gather additional information. What should you do in this situation to make the best possible decision? 1. You will talk to a class teacher who agrees with the psychologist's assessment and ask her for an opinion. You will also ask the psychologist to explain to you in more detail why she thinks your child is gifted and to describe the methods, conditions, and the price of the proposed program. 2. You will search the Internet to find the experiences of parents whose children are gifted and of parents whose children attended the proposed program at your school. You will explore the content and dynamics of the program and how it differs from the regular curriculum. 3. Although you are glad that the psychologist recognized your child's giftedness, you are not completely sure of her assessment. You want to make sure it's correct before you put your child in such a demanding program. You will take your child to the Center for Gifted Children for further evaluation.
8.	After a routine check-up, the doctor approaches you with a feedback. She informs you that you are generally in a good health but that, among other findings, she identified a potential health concern. While a relatively routine and straightforward surgical procedure could ensure that the issue doesn't become serious, the doctor assures you that it won't be necessary and there's no need for concern. As you have never noticed any symptoms or had any issues related to the concerns described, you also feel that excessive worry might not be warranted. However, you still want to gather additional information. What should you do in this situation to make the best possible decision?

(Continued)

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Table A1. *(Continued).*

No.	Item
	1. You will ask the doctor to provide you with a more detailed explanation of why they believe the surgical procedure is unnecessary. You are confident that this will give you a clearer understanding of what to do and help you feel more at ease. 2. You will request more information from the doctor about the specific health concern mentioned, and you will inquire about the experiences of individuals who have undergone the surgical procedure as well as those who have not. 3. You will ask the doctor to provide you with a more detailed description of this health concern, its symptoms, and long-term prognosis. In doing so, you will specifically ask her to outline the risks associated with not undergoing the surgery.
9.	You are a manager who has just been urgently hired by a company for web sales. You are employed as the head of the 'Procurement and Delivery System Management' and the CEO has immediately made it clear that they expect an efficient response from you to enhance the procurement segment of the system. According to the CEO's opinion, this segment is causing issues with orders and delays in delivering the ordered products. What should you do in this situation to make the best possible decision? 1. You are expected to fix things and to do so as quickly as possible. You dive into the work immediately, without too many questions. You will send out assignments to people to restructure the procurement system as soon as possible. 2. Before delving into changes, you will first have a conversation with the CEO and ask them to further explain the arguments for changing the procurement system. You want to have a clearer picture of what the CEO wants so that you don't miss potentially important information. 3. Even at the risk of appearing indecisive in the eyes of the CEO, you will hold off on concrete actions until you consult with other people in the organization who are well-versed in all of its business operations. You want to investigate whether the issue truly lies in the procurement system or if the delays are occurring due to other parts of the system, such as delivery.
10.	You are an owner of an agency for adventure tourism that conducts a smaller portion of its bookings online and a larger portion through a storefront located in the city center. Your business performed very well last season, and now you are considering renting additional space in the city center to further increase revenue. You are confident that this move would double your sales and significantly boost profits. However, the rent for a space in the city center is quite expensive (around $15,000 per month), and if it fails, it could potentially lead to financial difficulties. Currently, you have 15 employees in the company, most of whom have expressed support for the idea, while a few of the most experienced ones are seriously reserved. Additionally, you have an acquaintance in a similar industry who has already taken a similar step and is convinced it's the right move. What should you do in this situation to make the best possible decision? 1. You are convinced that renting the additional space is the right decision, so you will independently decide to proceed with the rental. You believe it doesn't make sense to overly delay the decision. 2. You will consult with your acquaintance who has already rented additional business space and considers it a good idea and ask them to explain why they made such a decision.

(Continued)

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Table A1. *(Continued).*

No.	Item
	3. You are aware that some of your employees disagree with the decision to rent the new space. Before making a decision, you will ask them to explain their viewpoints, and you are particularly interested in the arguments from employees who are skeptical about renting a new space.
11.	You have recently been promoted to the position of Human Resources Department Manager in a large company. The management expects you to introduce some changes to enhance employee motivation. An older colleague, a long-time employee of the HR department, believes from practice and experience that rewarding employees based on performance is the best way to motivate them. This approach also sounds like a very good move to you—it makes sense that people will exert more effort if they are paid in proportion to the work they have done, and you do not currently see any objective drawbacks to this method. What should you do in this situation to make the best possible decision? 1. You don't see any significant drawbacks to this approach, so you will introduce a performance-based reward system as soon as possible. This will increase employee motivation and simultaneously meet the management's demands. 2. You will have a conversation with the older colleague who advocates for this system and knows more about it than you do. If his reasons for implementing this approach are compelling enough, you will implement it right away. 3. Although it seems that the majority supports this approach, before making a decision about implementation, you will make every effort to find key arguments against this approach, or identify potential issues with introducing this practice.
12.	Local elections for the mayor will soon be held in your city. Your favorite is candidate A. You watched an interview with him and you believe he presented convincing arguments in support of the measures he is advocating. His opponent, candidate B, is a man whose worldview is completely different from yours, so you didn't even bother to listen to his arguments about local problems. You have decided to vote for candidate A. On the day before the election, an apolitical friend asks you for your opinion on which candidate is better and who he should vote for. Your friend sees you as someone who is much more knowledgeable and familiar with the local political scene than he is. What should you do in this situation? 1. Tell your friend to vote for candidate A, without mentioning that you know next to nothing about candidate B's program. No one has all the information and you trust your own judgment. 2. Indicate that you are not completely familiar with candidate B's program, but explain to your friend why you will still vote for candidate A. 3. Openly tell a friend that you will vote for candidate A, but that you don't know enough about candidate B to make a sound judgment and give him an advice.
13.	You are a red wine enthusiast. For many years, you have been learning about wines, and you're a favorite among your friends because you always know how to choose a good wine and describe it in a way others don't. You regularly consume it—you have a glass every day after lunch or dinner, and you're happy because you've read the studies that wine is good for the cardiovascular system. In a conversation with a friend, you learn about the latest research he read about, which claims that even in the smallest amount red wine does not contribute to health. On the contrary, the research shows that all types of alcohol, including wine, are ultimately harmful. What should you do in this situation to make the best possible decision?

(Continued)

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Table A1. *(Continued).*

No.	Item
	1. You will disregard this information. The contents of internet portals are often inaccurate, and the new research is likely questionable, given the conflicting results of all previous studies. You will explain to your friend that he shouldn't trust such information. 2. You are willing to listen to the new research. While you believe in the accuracy of previous information, there's always a chance that previous studies were wrong. However, you still believe that wine can contribute to health when you don't overconsume it. 3. After hearing about the new research, you are no longer as certain about the benefits of red wine. You know that newer research can uncover something that previous studies missed. You admit to your friend that there's a fair chance he might be right.
14.	You are making a holiday lunch that you have been preparing for years using the same recipe. It is one of your favorite dishes and you have received many compliments for it. Everyone who knows you know that you are the master of that dish. Your mother-in-law is visiting and advises you to add additional spice to your dish to make it even tastier. She prepares it in a slightly different way and emphasizes that everyone who has tried it likes it very much. What should you do in this situation? 1. You will explain that you have been preparing the dish according to the same recipe for years and that it does not need additional spices. You will completely reject her suggestion. 2. You'll thank her for the advice and admit that the spice could improve the dish. However, you will emphasize that this is an important holiday lunch, so you don't want to experiment. 3. You will admit that you hadn't thought about this spice before and that you can see how it could really enrich your dish and that it would be worth trying.
15.	You are an architect, and in your 15-year-long career, you have designed several award-winning and high-quality residential building. You are having a conversation with a statics expert about the construction of a new building. You have invested a lot of effort and time considering different aspects of the design, including a certain type of staircase that you think should be installed in the building. However, the statics expert tells you that this type of staircase will greatly reduce the static resistance of the building and should be changed. However, changing the staircase would lead to numerous additional changes to the design. What should you do in this situation to make the best possible decision? 1. You should tell the expert that you have thoroughly checked all the possibilities and that this option is the best if we take into account the other components of the design. You refuse to change your plan and stand by your decision, trusting your judgment. 2. You say that you thought about it and generally agree with the expert's comment. After additional argumentation, you stand by your initial position, emphasizing your vast experience. 3. You will discuss the matter with an expert and admit that in this case he may be more knowledgeable than you and that your initial conclusion might have been wrong.

(Continued)

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

Table A1. *(Continued).*

No.	Item
16.	You are the manager at a marketing agency with 20 employees. The director of a hotel contacted you and wants to hire you to 'rebrand' the hotel as a great destination for young families. The problem is that the tourist season starts in a month, and the hotel delayed this until the last hour due to financial uncertainties. However, the director decided on a more aggressive campaign and it is extremely important to him that it be successful. You are the most experienced member of the agency, and you have been in several situations in your career when you had to do something at the last minute. Based on these experiences, you have a clear idea of how you would implement this campaign and you feel that, due to the short deadline, it is best to just communicate your idea to the team and divide them into roles. However, not everything goes smoothly at the team meeting. A junior colleague comments on your plan and shows disagreement with it, persistently trying to push his idea. Although his arguments make some sense, the deadline for a decision is relatively short. What should you do in this situation? 1. React vigorously and make it clear that you do not want to listen to further complaints. You have nothing against other people's ideas, but your rich experience gives you the right to persist in your idea. In addition, you will not allow a junior colleague to 'undermine' your authority. 2. You will thank your colleague for fresh ideas and say that you appreciate her engagement and creativity. You will admit that her remarks make sense, but that the current situation requires a quick reaction and a great experience, so this time you will still 'play it' your way. 3. You will admit that your colleague made you reconsider your decision. In addition, you will also encourage the rest of the team not to hesitate to criticize your approach if they think you are wrong, even at risk of having your authority questioned.

https://doi.org/10.1017/jdm.2025.10008 Published online by Cambridge University Press

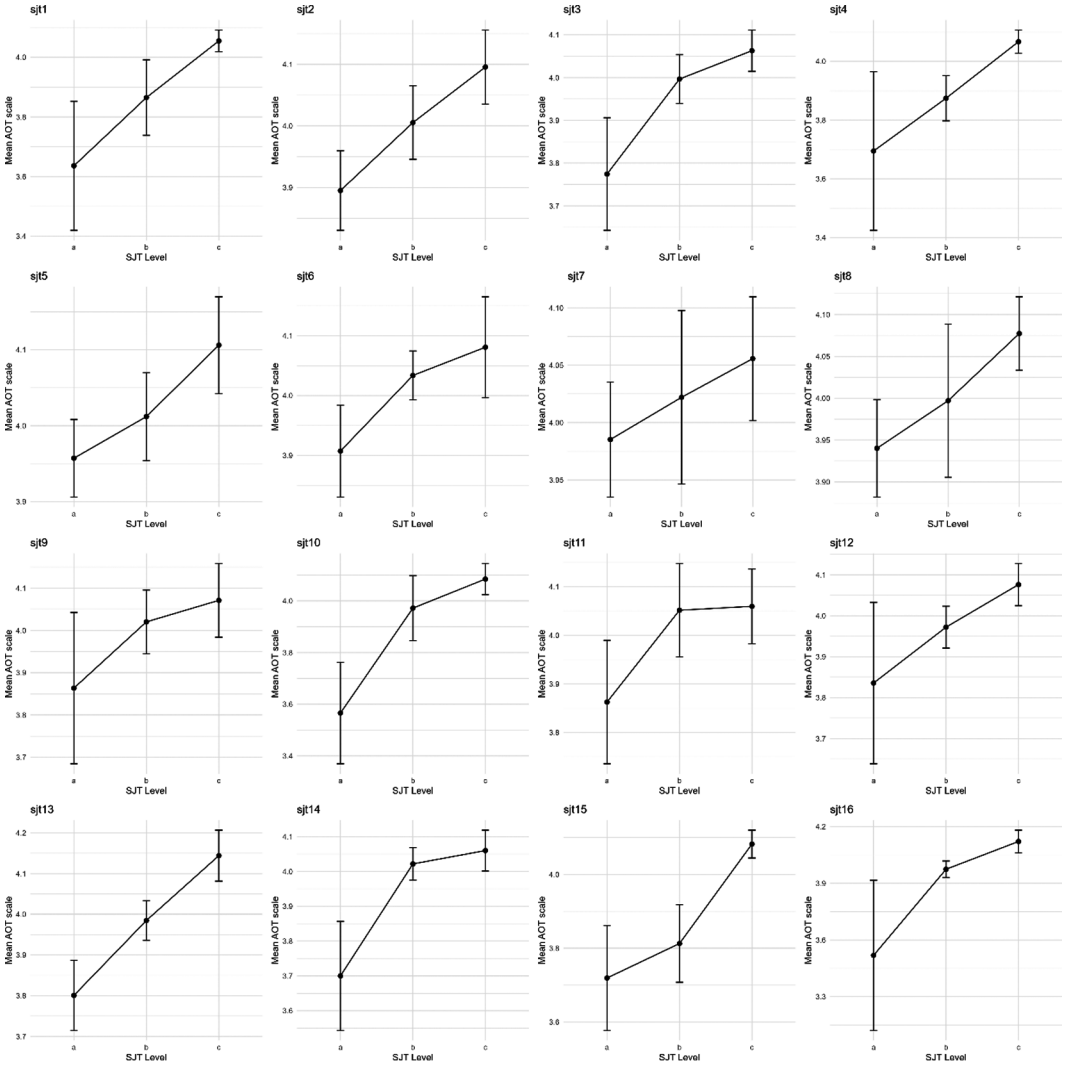

Figure A1. *Average scores on the AOT scale for participants choosing different response options on AOT-SJT, for each of the 16 AOT-SJT items (option 'c' indicates the highest AOT response in SJT items, whereas option 'a' indicates the lowest).*

Cite this article: Erceg, N., Vrhovnik, A., Galić, Z., and Ružojčić, M. (2025). Development and initial validation of a situational judgment test for the measurement of actively open-minded thinking. *Judgment and Decision Making*, e32. https://doi.org/10.1017/jdm.2025.10008