Sensors and Materials, Vol. 38, No. 3 (2026) 1303–1318
MYU Tokyo

S & M 4378

A Fundamental Study of Introspective Sensing Framework Combining Ambiguous Responses with Decision Latency

Takahiro Ueno,[1*] Mina Nakano,[2] and Noboru Nakamichi[1,3]

[1]Faculty of Engineering, Fukuyama University, 1 Sanzo, Gakuen-cho, Fukuyama, Hiroshima 729-0292, Japan
[2]Faculty of Human Cultures, Fukuyama University, 1 Sanzo, Gakuen-cho, Fukuyama, Hiroshima 729-0292, Japan
[3]ANKR DESIGN Inc., 4-4-17 Minamishinagawa, Shinagawa, Tokyo 140-0004, Japan

(Received November 27, 2025; accepted February 10, 2026)

Keywords: introspective sensing, uncertain option, decision latency, accelerometer, mental health tracking

Recent advances in wearable sensing and digital phenotyping have enabled mental health tracking by combining self-reports with physiological and behavioral indicators. However, conventional interfaces record only final answers and overlook how users hesitate and introspect. This gap contributes to discrepancies between self-reports and behavioral signals. In this study, we propose an introspective sensing framework that treats the response process itself as a sensing target. We combine an explicit uncertain option '?' with decision latency measured by a smartwatch accelerometer. Forty-one participants answered ten introspection-inducing questions about personality-related tendencies on a smartwatch interface offering three symbolic response options: '◯' (yes), '?', and '×' (no). Experimental results showed that 61% of participants used '?' at least once. Decision latency was longest for '?', followed by '×' and '◯'. When responses were grouped into three types (immediate, internalized, and externalized) based on latency and '?' usage, each type exhibited distinct decision patterns. These patterns suggest different balances of intuitive versus deliberative processing and different ways of expressing uncertainty. Our findings indicate that ambiguous responses and decision latency provide behavioral features that complement physiological indicators and support low-burden, stigma-sensitive mental health tracking.

1. Introduction

Recent advances in IoT technology and wearable sensing have enabled the continuous tracking of physiological and behavioral data in daily life. Ecological momentary assessment (EMA) presents question prompts to users at random or scheduled times on smartphones or smartwatches, reducing recall bias and enabling the fine-grained characterization of daily patterns.(1,2) Digital phenotyping combines self-reports with passively sensed data from smartphones and connected devices to monitor mental health states in situ.(3,4) Standard clinical scales such as PHQ-8/9 are repeatedly administered via mobile apps, while passive sensing

*Corresponding author: e-mail: t-ueno@fukuyama-u.ac.jp
https://doi.org/10.18494/SAM6085

ISSN 0914-4935

continuously records activity, mobility, and sleep patterns.(5,6) Most self-report interfaces are direct digital implementations of conventional questionnaires. Users rate their current mood, stress, or symptoms on numeric Likert scales, visual analogue sliders, or categorical option.(1,7,8) Users select exactly one category with a tap or swipe, and the system records only the final response and timestamp. The underlying design assumption is that a single "correct" category represents the user's current state.

However, recent digital phenotyping studies have repeatedly reported discrepancies between self-reported symptoms and passively sensed behavioral or physiological indicators. These include weak correlations between EMA-based stress reports and wearable-derived indicators,(9) the limited explanatory power of smartphone-sensed behavioral features beyond self-reports,(5,10) and modest performance gains from passive sensing for suicidal ideation and behaviors,(11) indicating that simple one-to-one mappings from behavioral indicators to self-reported mental health states are insufficient.

One important source of this discrepancy lies in the introspection process during self-reporting. When asked "Are you feeling stressed right now?", users must recognize their internal state and map it onto the provided response options. In depressive states, the awareness of one's own condition can be blunted, making clear judgments difficult.(12) Users may require additional reflection time or answer based on habitual patterns rather than momentary experience. Current mental health tracking interfaces rarely capture this introspection process. Systems employing Likert scales, sliders, or emoji selections do not provide an explicit option for "I am not sure about my state," forcing users to select polarized answers even when their internal state is ambiguous. This design creates problems such as polarized response styles toward either "midpoint" or "extremes."(13) The objective estimation of mental health states using physiological indicators also has limitations. Reliable estimation often requires several minutes of measurement time, making it difficult to capture psychological states during brief EMA prompts. Furthermore, physical activities introduce noise into physiological signals and degrade estimation accuracy. Consequently, physiological and behavioral indicators alone cannot fully explain momentary self-reports.

In this study, we address this challenge by proposing an introspective sensing framework that treats the introspection process itself as a sensing target. We combine two key components. The first is an explicit uncertain option '?' alongside conventional '◯' (yes) and '×' (no) options, acknowledging "I cannot judge" as a legitimate response. The second is the decision latency measured using a smartwatch accelerometer, capturing subtle wrist movements, stationary periods, and hesitation patterns during the response process. By integrating explicit ambiguous responses with decision latency measurement, the proposed framework enables recording not only "what was answered" but also "how the user introspected and arrived at the answer." This framework provides two advantages. First, the system can distinguish between immediate intuitive decisions and decisions made after extended reflection. Second, behavioral indicators from the response process can be combined with physiological indicators to help interpret discrepancies between self-reports and objectively measured signals.

In this present study, we evaluate the basic properties of the proposed introspective sensing framework, addressing the following research questions (RQs):

- RQ1: To what extent is the explicit uncertain option utilized in mental health tracking, and what response patterns does it reveal?
- RQ2: What differences in response characteristics emerge when participants are categorized on the basis of decision latency and ambiguous response usage patterns?

We conducted smartwatch-based experiments in which participants answered introspection-eliciting questions using the proposed interface, and analyzed the relationship between ambiguous response usage and decision latency to evaluate the validity of the introspective sensing framework.

2. Proposed Framework

Traditional mental health tracking systems overlook uncertainty and hesitation in the response process. To address this limitation, we propose an introspective sensing framework that combines an uncertain option '?' with decision latency measurement. Figure 1 shows an overview of the proposed framework. Our framework provides three response options for mental health questions such as "Do you feel stressed?". In addition to the conventional options '○' (yes) and '×' (no), we introduce an explicit uncertain option '?'. The response procedure consists of two steps. First, users lower their arms and wait. Second, after receiving a vibration notification, users raise their wrists and respond to the question. During this process, the system measures the stillness duration of the wrist using an accelerometer sensor. This duration is recorded as decision latency.

The proposed framework uses a smartwatch platform. This allows users to perform introspection through the natural action of checking the current time on their watch. The system captures the introspective process from this action. Conventional interface design demands clear-cut answers. In contrast, our explicit '?' option enables users to express uncertainty concretely. The framework is grounded in two theoretical frameworks. The first framework views ambiguous expressions as cognitive resources. The second framework is dual-process theory. The following subsections explain these theoretical foundations and describe response types based on the proposed framework.

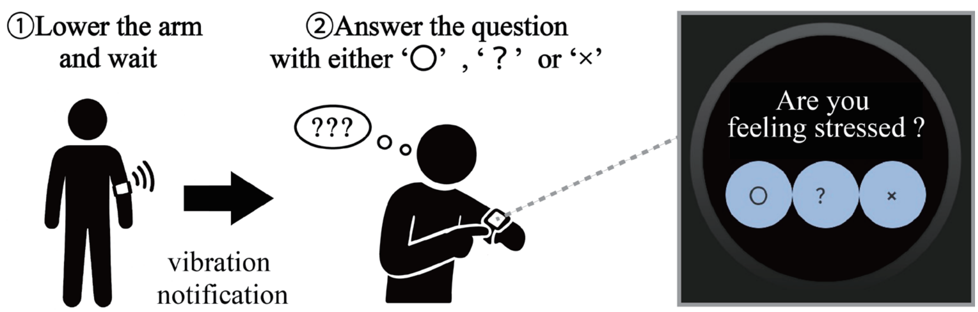

Fig. 1. (Color online) Introspective sensing framework combining the uncertain option '?' with decision-latency measurement.

2.1 Ambiguous expression and promotion of introspection

Traditionally, ambiguity has been positioned as a factor that impairs usability. Therefore, ambiguity has been treated as something to be eliminated in interface design. However, Gaver *et al.* reevaluated ambiguity as a design resource that encourages users' active interpretation and deep engagement.(14) Ambiguity arises during the process in which users discover meaning in an interface based on their own context and perspective.(14) It is not mere obscurity but functions as interpretive space.(15)

The value of providing an explicit ambiguous response option fundamentally differs from the "midpoint" in Likert scales. The midpoint indicating "neither" suggests the neutrality of attitude. In contrast, the uncertain option '?' has a different meaning. This option explicitly represents cognitive states such as "cannot judge" or "not confident." This distinction enables our framework to capture user uncertainty independently from other response patterns. Previous research demonstrates this principle. Sanches *et al.* developed "Affective Health."(16) This system presents biosignals through abstract and ambiguous visuals. As a result, users proactively interpret their fluctuating states. This process promotes personal introspection and diverse interpretations. The uncertain option '?' has important significance. This option acknowledges that "I don't know my own state" as legitimate. Furthermore, it treats this uncertainty as a valid response. Conventional interfaces force users to make clear choices. Users must select either '◯' or '×', or choose a number from 1 to 5. Consequently, users are compelled to artificially clarify ambiguous states. In contrast, our framework introduces the uncertain option '?'. This option enables users to express uncertainty in their judgment process as it is. As a result, more authentic self-reporting is supported.

2.2 Dual-process theory and decision latency

Dual-process theory is a framework that explains human decision-making.(17,18) This theory proposes two cognitive processing systems (Fig. 2). System 1 performs fast and intuitive

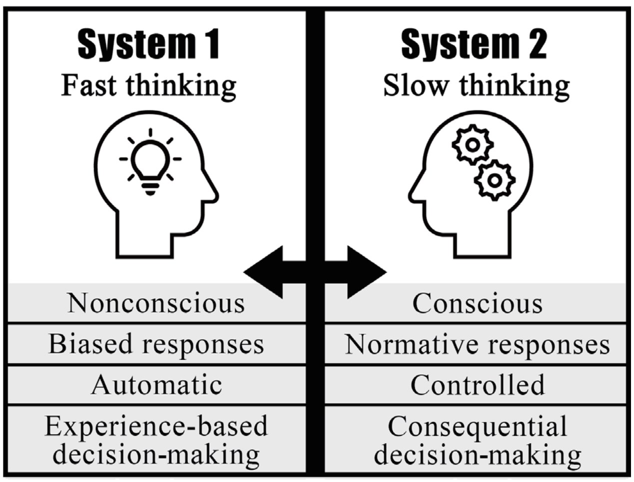

Fig. 2. Concept of dual-process theory by systems 1 and 2.

processing. System 2 performs slow and deliberative processing. Decision latency during responses serves as an indicator to distinguish these two systems. Decision latency shows which system is operating. However, decision latency alone has limitations and cannot identify "why the response is slow." Long decision latency may reflect two different situations. The first situation is when users carefully compare and evaluate multiple options. The second situation is when users experience hesitation based on uncertainty. In this case, users feel "I don't really know, but I must choose something." Conventional interfaces could not distinguish between these two situations. Our framework enables this distinction. The proposed approach combines the uncertain option '?' with decision latency measurement. Previous research has demonstrated the effectiveness of decision latency. In Chung *et al.*'s study, a relationship between depression severity and response latency was confirmed.(19) This shows that decision latency serves as a valid indicator of psychological states.

2.3 Classification of response types

We propose a framework that classifies participants into three response types. This classification is based on decision latency and ambiguous response usage patterns. The three response types are immediate, internalized, and externalized.

- Immediate type consistently shows short decision latency. Users of this type select symbols quickly. This response pattern reflects the intuitive judgment of System 1. The immediate type shows response characteristics without hesitation regarding question content.
- Internalized type shows long decision latency. However, this type does not use the uncertain option '?'. The internalized type always selects clear polar options ('◯' or '×'). This response pattern reflects deliberative processing of System 2. However, the internalized type does not express internal hesitation externally.
- Externalized type has two characteristics. First, this type actively uses the uncertain option '?'. Second, the externalized type shows long decision latency. This response pattern reflects the deliberative processing of System 2. The externalized type responds while considering uncertainty. This type explicitly expresses uncertain choices.

Conventional interface design has limitations. These interfaces cannot distinguish between internalized and externalized types. Both are classified into the same category as "slow responses." However, our framework can distinguish these two different cognitive characteristics by combining the uncertain option with decision latency measurement. To evaluate the validity of this introspective sensing framework, we conducted a smartwatch-based experiment examining the relationship between ambiguous response usage and decision latency patterns across these response types.

3. Experiment

We investigated the relationship between ambiguous responses and decision latency in the proposed framework. The experiment employed a question-answering task using a smartwatch. Forty-one healthy Japanese participants took part in the study (mean age = 28.5 ± 13.0 years, 36 males, 5 females). All participants were right-handed and had normal or corrected-to-normal vision. This study was approved by the Research Safety and Ethics Committee of Fukuyama University (Approval number: 2025-H-36). Informed consent was obtained from all participants through both verbal explanation and on-device confirmation on the smartwatch before participation.

3.1 Experimental procedure

Figure 3 shows the experimental procedure. Participants were instructed to keep their arms lowered between trials. After a waiting period of several seconds, the smartwatch issued a vibration notification. Participants then raised their wrists to check the display and viewed the presented question. The display showed one personality-related question and three response symbols: '◯' (yes), '?', and '×' (no). When participants selected '?', they additionally indicated whether their response leaned toward '◯' or '×'. These responses were recorded as '?◯' or '?×'. In statistical analysis, the '?' response category encompasses both '?◯' and '?×' subcategories. Note that the classification into '?◯' and '?×' was implemented for experimental analysis purposes and is not included in the practical system.

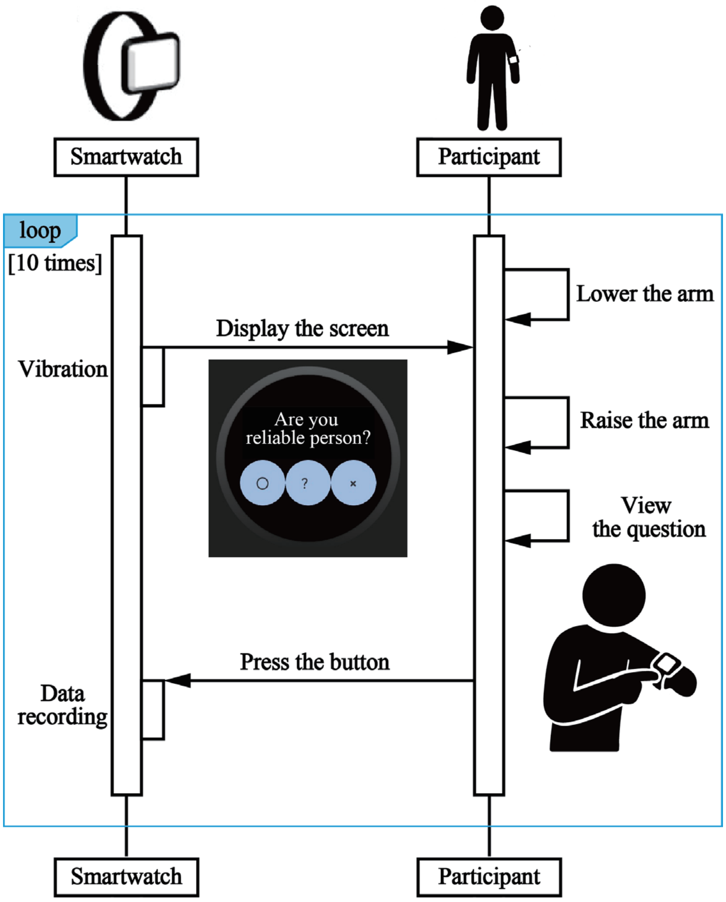

Fig. 3. (Color online) Sequence diagram of experimental procedure.

Table 1 shows the questions used in this experiment. These ten items were designed on the basis of the Big Five personality framework.[20] The Big Five comprises five personality traits: extraversion, agreeableness, conscientiousness, neuroticism, and openness. Questions directly asking about mood or emotion tend to elicit positively biased responses due to social desirability effects, making it difficult to collect diverse response patterns. Therefore, we employed personality-related questions in this study. Questions about personality traits are inherently abstract, requiring participants to reflect on their self-perception. This design encourages introspection. For each personality trait, we designed one question with positive framing and one with negative framing (e.g., for extraversion: “Are you an active person?” and “Are you a reserved person?”). When participants have a clear self-perception of their personality traits, they would show contrasting response patterns of ‘◯’ and ‘×’ for each pair. This design minimizes the psychological barrier to selecting each symbol. The experiment was conducted in an indoor environment with minimized external interference. Some participants used only three response options (‘◯’, ‘?’, and ‘×’) without the additional question about ‘?◯’ and ‘?×’.

3.2 Measurement of decision latency

We used a Google Pixel Watch 3 to measure decision latency. Accelerometer data were collected at a sampling frequency of approximately 16 Hz.[21] Figure 4 shows an example of time-series accelerometer data during a ‘?’ response. The question response time represents the period from the vibration notification to the participant’s symbol selection. In Fig. 4, the accelerometer values begin to increase after some delay following the vibration, indicating the initiation of the wrist-raising movement. Before the ‘?’ response, the sensor values remain close to zero, indicating that the wrist is stationary. We defined this stationary period as decision latency. Based on the characteristics of the response movement, we divided the question response time into two phases:

(1) Wrist-raising movement after vibration
(2) Stationary period during response selection

From the accelerometer data, we calculated the amount of acceleration change Δ_t at each time point using the following equation:

Table 1
Personality questions based on Big Five framework.

Personality trait	Question
Extraversion	Are you an active person? Are you a reserved person?
Agreeableness	Are you a selfish person? Are you considerate of others?
Conscientiousness	Are you a reliable person? Are you a careless person?
Neuroticism	Are you a calm person? Are you a worrier?
Openness	Do you like new things? Do you have trouble coming up with ideas?

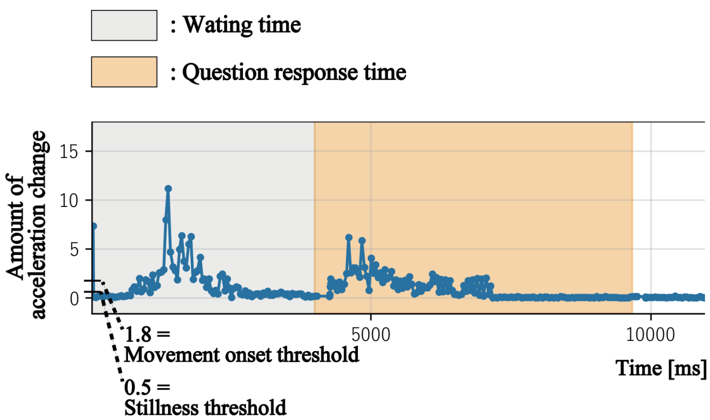

Fig. 4. (Color online) Example of time-series accelerometer data during ‘?’ response.

$$\Delta_t = \sqrt{\left(x_t - x_{t-1}\right)^2 + \left(y_t - y_{t-1}\right)^2 + \left(z_t - z_{t-1}\right)^2}, \tag{1}$$

where x_t, y_t, and z_t are the three-axis accelerometer values [m/s^2] at time t. We defined decision latency as the cumulative duration of stationary intervals occurring between the vibration notification and response selection. Stationary intervals were identified as periods satisfying Δ_t =< 0.5 m/s^2. To distinguish between movement and stationary states, we applied a movement onset threshold of 1.8 m/s^2 and a stationary threshold of 0.5 m/s^2.

Let K denote the total number of stationary intervals detected in a single trial. Each stationary interval i consists of m_i consecutive samples. The duration d_i of this interval is calculated as

$$d_i = m_i \times \left(1/f\right), \tag{2}$$

where f is the sampling frequency. The decision latency DL for each trial is defined as the sum of all stationary interval durations:

$$DL = \sum\nolimits_{i=1}^{K} d_i. \tag{3}$$

For each participant, we calculated the mean decision latency $\overline{DL}$ for each response symbol because individual differences exist in decision latency. The mean decision latency $\overline{DL}_{j,s}$ for participant j when selecting symbol s ∈ {‘◯’, ‘×’, ‘?’, ?◯‘’, ‘?×’} across all trials is defined as

$$\overline{DL}_{j,s} = \frac{1}{n_{j,s}} \sum\nolimits_{i=1}^{n_{j,s}} DL_{j,s,i}, \tag{4}$$

where $DL_{j,s,i}$ is the decision latency for the i-th trial in which participant j selected symbol s and $n_{j,s}$ is the total number of such trials.

3.3 Classification and analysis of response types

We classified participants' response types on the basis of the framework described in Sect. 2.3. The classification used mean decision latency, standard deviation, and the frequency of '?' option usage. Figure 5 shows the classification procedure.

- **Immediate type:** Both mean decision latency and standard deviation are low. This indicates rapid and consistent responses based on System 1 processing.
- **Internalized type:** Mean decision latency is long but '?' is not used. This indicates internalized hesitation during System 2 processing.
- **Externalized type:** The '?' option is actively used. This indicates the explicit externalization of uncertainty during System 2 processing.

For statistical analysis, we applied the Friedman test when comparing $\overline{DL}$ across multiple symbols ('◯', '?', and '×') within participants. For the comparison of $\overline{DL}$ between two symbols, we used the Wilcoxon signed-rank test. The significance level was set at $p < 0.05$ for all tests.

4. Results

4.1 Overall response tendencies by symbol

Forty-one people participated in this study (N = 41). Each participant responded to 10 questions. The dataset consisted of 410 samples (S = 410), each containing symbol selection and accelerometer-based decision latency measurements. Table 2 presents the statistical results of

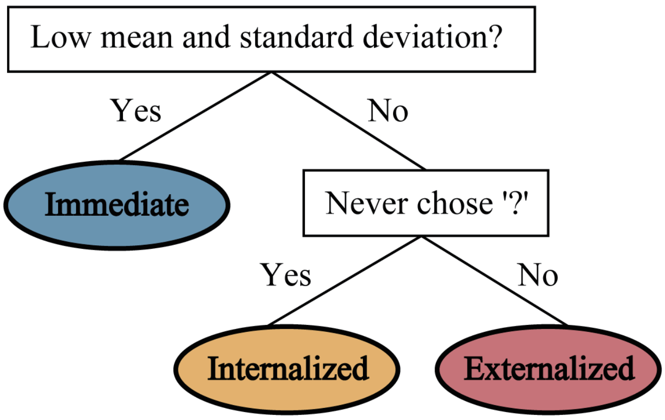

Fig. 5. (Color online) Procedure for classifying response types.

Table 2
DL statistics by response symbol.

Symbol	DL (Mean ± std) (ms)	N	S
◯	1295 ± 798	41	232
?	2352 ± 1973	25	68
×	1680 ± 1035	40	110
?◯	1663 ± 1079	13	29
?×	2782 ± 2899	13	17

decision latency for each response symbol. Among all 410 responses, '◯' was selected 232 times (56.6%), making it the most frequent choice. '×' was selected 110 times (26.8%) and '?' was selected 68 times (16.6%). This result indicates that participants tended to prefer positive responses. Twenty-five participants (61.0%) used '?' at least once. *DL* varied systematically across symbols. '?' showed the longest *DL* (2352 ± 1973 ms), followed by '×' (1680 ± 1035 ms) and '◯' (1295 ± 798 ms). This consistent ordering ('?' > '×' > '◯') indicates that ambiguous responses required the longest deliberation, while positive responses were made most quickly. The comparison of mean *DL* differences revealed that '?' was 1057 ms longer than '◯' and 672 ms longer than '×'. '×' was 385 ms longer than '◯'. This result suggests that negative judgments involved a more deliberation than positive judgments. We analyzed cases where participants selected '?' and subsequently polarized their choice. The mean *DL* for '?◯' was 1663 ± 1079 ms, while '?×' was 2782 ± 2899 ms. The difference between the two was 1119 ms.

4.2 Classification of response types

We classified participants into three response types following the procedure shown in Fig. 5. Table 3 presents participant rankings based on mean decision latency and standard deviation. First, we identified participants who consistently showed short decision latencies. We extracted the top 25% of participants for both mean *DL* and standard deviation rankings. This top 25% corresponded to approximately 10 participants. Eight participants appeared in both rankings (IDs 16, 32, 20, 27, 14, 13, 28, and 40). The mean *DL* of these eight participants ranged from 220 to 970 ms. Their standard deviations ranged from 181 to 682 ms. These results indicate consistently fast and stable responses. Therefore, we classified these participants as the immediate type. We then analyzed the remaining 33 participants. Among them, 13 participants never used '?'. We classified these participants as the internalized type. The remaining 20 participants used '?' at least once. We classified these participants as the externalized type.

4.3 Decision latency by symbol across response types

Table 4 shows $\overline{DL}$ for each symbol across the three response types. Figure 6 presents the results as boxplots. Table 5 shows statistical testing results for each response type.

Table 3
DL statistic ranking of participants.

Rank	Participant ID	Mean (ms)	Participant ID	Std (ms)
1	16	220	16	181
2	32	263	32	231
3	20	434	20	231
4	27	587	22	470
5	14	620	11	501
6	13	765	18	525
7	28	844	31	645
8	40	912	27	649
9	22	915	13	675
10	18	970	14	682

Table 4
$\overline{DL}$ statistics by response type and symbol.

Type		Symbol				
		◯	?	×	?◯	?×
Immediate	$\overline{DL}$	518 ± 281	842 ± 303	624 ± 527	839 ± 372	832 ± 277
	N	8	5	8	3	4
Internalized	$\overline{DL}$	1655 ± 606	—	1750 ± 913	—	—
	N	13	—	13	—	—
Externalized	$\overline{DL}$	1371 ± 854	2729 ± 2036	2077 ± 996	1910 ± 1101	3650 ± 3135
	N	20	20	19	10	11

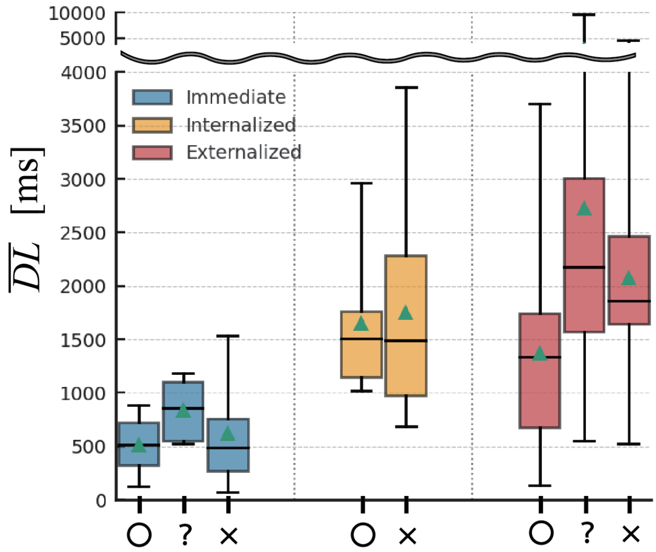

Fig. 6. (Color online) Boxplots of decision latency by response type and symbol.

Table 5
Statistical testing results by response type.

Type	Comparison	N	p-value
Immediate	◯, ×, ? groups	5	0.549
	◯ vs ×	5	0.813
	◯ vs ?	5	0.125
	? vs ×	5	0.813
Internalized	◯ vs ×	13	0.735
Externalized	◯, ×, ? groups	19	0.005
	◯ vs ×	19	0.040
	◯ vs ?	20	0.007
	? vs ×	19	0.156

- **Immediate type (8 participants):** All symbols had *DL* values below 900 ms. The ordering of *DL* was '◯' < '×' < '?'. However, the mean difference between '◯' and '?' was small at 324 ms. Statistical analysis revealed no significant differences among symbols (p = 0.549). Five of the eight immediate-type participants selected '?' at least once. '?◯' (839 ± 372 ms) and

'?×'(832 ± 277 ms) showed almost no difference. These results indicate that the Immediate type responded intuitively to all symbols.

- **Internalized type (13 participants):** This type showed longer decision latencies than the immediate type. The mean difference between '◯' and '×' was 95 ms, with no significant difference ($p = 0.735$).
- **Externalized type (20 participants):** This type showed a clear hierarchical structure in *DL*. '◯' was 1371 ± 854 ms, '×' was 2077 ± 996 ms, and '?' was 2729 ± 2036 ms. '?×' was 1740 ms longer than '?◯'. The Friedman test revealed significant differences among the three symbols ($p = 0.005$). We conducted post-hoc comparisons using the Wilcoxon signed-rank test. Both '×' ($p = 0.040$) and '?' ($p = 0.007$) were significantly longer than '◯'. No significant difference was found between '?' and '×' ($p = 0.156$).

5. Discussion

5.1 Characteristics of three response types

In this study, we defined the following three response types based on participants' response patterns and decision latency: immediate, internalized, and externalized.

Immediate type consistently exhibits short decision latency across all response symbols. This type reflects System 1 in dual-process theory. System 1 is characterized by fast and intuitive decision-making. The immediate type is considered to represent a stable cognitive style under normal conditions. However, if a transition from the immediate type to the internalized or externalized type is observed, this may serve as an early indicator of new stressors or cognitive overload.

Internalized type does not use the '?' option. This type exhibits long decision latency for both '◯' and '×' ('◯': 1655 ± 606 ms; '×': 1750 ± 913 ms). No significant difference was found in decision latency between '◯' and '×'. This result indicates that the psychological barriers for both responses are approximately equal. The internalized type does not consider '?' as a choice option. This type is considered to engage in cognitive processing based on binary judgment. However, if the balance of response patterns between '◯' and '×' becomes disrupted, or if decision latency becomes extremely prolonged, these may indicate signs of psychological burden or internal conflict.

Externalized type actively utilizes the '?' option. This type showed a clear hierarchical structure in decision latency ('◯' (1371 ± 854 ms) < '×' (2077 ± 996 ms) < '?'; $p = 0.005$). The decision latency of this type associated with '?' responses was notably longer than those of the other types. Specifically, '×' (3650 ± 3135 ms) showed a longer decision latency than '◯' (1910 ± 1101 ms). Furthermore, a difference in decision latency was observed between '◯' and '×'. This result suggests that conflict in hesitating between '×' and '?' persists in the externalized type. '?' is not merely an intermediate choice option. This response option functions as a means of externalizing psychological resistance associated with negation. Explicit ambiguous responses induce introspection. Decision latency is an effective indicator for quantitatively capturing this introspective process.

5.2 Applicability to mental health tracking

The proposed framework combines ambiguous responses with decision latency. This framework can measure introspective processes as behavioral indicators, which conventional mental health tracking systems could not capture. The explicit ambiguous response '?' plays an important role. This response functions as a means of expressing judgment difficulty itself, rather than being merely an intermediate choice. This study demonstrated that three response types (immediate, internalized, and externalized) can be identified. By tracking temporal changes in these types, it may be possible to detect early changes in psychological states.

The behavioral indicators obtained through this framework can serve as supplementary information for interpreting discrepancies between self-reports and physiological indicators. Physiological indicators have difficulty reflecting short-term hesitation and introspection. Therefore, inconsistencies with self-reports often become problematic. This framework may address this issue. For example, consider an externalized-type participant who shows normal physiological indicators. However, if they simultaneously exhibit long decision latency for '?', this may indicate signs of psychological distress that are difficult to capture through physical responses. The opposite pattern is also possible. If an internalized type shows physiological tension yet immediately selects '◯', this may suggest that perceived distress is being underreported. Response types and decision latency enable more multifaceted mental health assessment.

This experiment used questions about personality traits. However, the same framework can be applied to other questions. Application to questions about social motivation and behavioral willingness is possible, for example, applying this framework to questions about behavioral intentions such as "Do you want to go to work today?" In this case, through increases in ambiguous responses or decision latency, it may be possible to capture early signs of social withdrawal tendencies or reduced affect. Such application examples demonstrate that this framework has extensibility in that it targets judgment processes as sensing objects.

This study has several limitations. First, we set thresholds for stillness detection from accelerometer data. We also defined classification criteria for response types. These were determined heuristically. Moreover, the accelerometer-based stillness measure can be affected by body movement and individual differences (e.g., tremor due to medical conditions), which may reduce robustness if a single set of thresholds is applied to all users. Future work should explore the personalization or calibration of thresholds. Validation using statistical methods is required in the future. Second, in this study, we adopted a specific interface design. Specifically, the structure consists of '◯': yes, '?': uncertain, and '×': no, representing positive polarity vs ambiguity vs negative polarity. Individual psychological characteristics may affect usage patterns of ambiguous responses. Social desirability, self-stigma, and perfectionism fall into this category. Analysis considering these factors will be necessary in the future. Third, this experiment was conducted in an indoor environment. Validation in daily life environments is essential. Furthermore, because the experiment included only ten questions in a short session, some participants may not have used ‘?’ simply because of limited opportunities. Longer and more frequent prompting in daily-life settings will be necessary to assess how often ‘?’ occurs in practice and whether response-type classification remains stable over time. In daily environments, environmental factors may affect response patterns. Effects of habituation through repeated use also need to be considered. The verification of the temporal stability of response types is also required.

6. Future Directions

Figure 7 illustrates the mental health tracking system using the proposed framework. In this study, we focused only on the fundamental component of this system: the relationship between ambiguous responses and decision latency obtained from smartwatch-based interaction. In practical use, however, the system is expected to operate as a continuous loop consisting of three stages: (1) the collection of mental health data on the smartwatch and transmission to the cloud, (2) the estimation of mental health status from accumulated time-series data, and (3) feedback to the user through euphemistic or indirect expressions.

In Stage (1), users simply respond to brief questions using the symbols ‘◯’, ‘?’, and ‘×’ while their decision latency is measured. Unlike conventional EMA interfaces that require users to explicitly recognize and rate their internal state, this system allows users to signal difficulty of judgment and hesitation through the ‘?’ option and prolonged decision latency. In that sense, symptom-like states can be reported without forcing users to clearly label or verbalize their own condition. This design resonates with prior work that treats ambiguity and abstract representations as resources for supporting reflection rather than usability problems.[14,16]

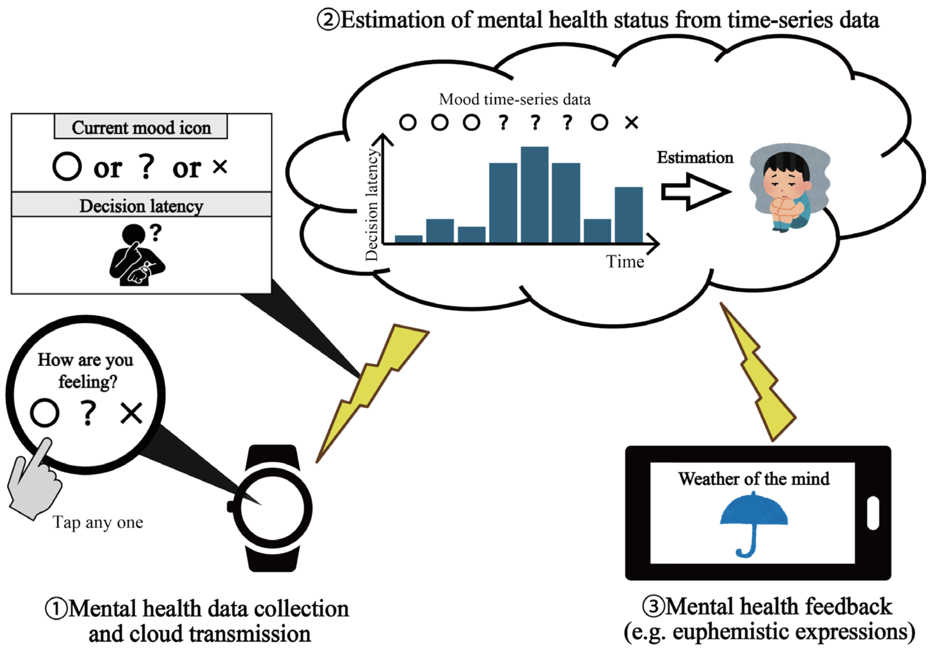

Fig. 7. (Color online) Mental health tracking system using the proposed framework.

Stage (2) extends the present analysis to longitudinal use. While the current experiment examined single-session responses, long-term deployment will enable tracking temporal changes in response types, decision latency, and the balance between '◯', '×', and '?'. Shifts from immediate to internalized or externalized patterns, or gradual increases in hesitation and negative-leaning ambiguous responses, may serve as early indicators of emerging stress, social withdrawal tendencies, or emotional flattening. The proposed framework thus provides a new class of behavioral features that complement physiological indicators in digital phenotyping.

In Stage (3), the estimated tendencies are fed back to users. Instead of presenting diagnostic labels or explicit risk scores, the system can use euphemistic expressions, such as weather metaphors, to gently convey changes in psychological state. Prior studies on visual metaphors and mental health campaigns suggest that such indirect messaging can reduce perceived stigma and improve engagement with mental health information.[22] Future work will experimentally compare direct versus euphemistic feedback conditions, measuring outcomes such as perceived stigma, self-stigma, reporting burden, and help-seeking intentions. Through these extensions, the proposed introspective sensing framework may evolve into a low-burden mental health tracking approach that captures subtle cognitive and emotional processes while mitigating the psychological costs of self-monitoring.

7. Conclusions

In this study, we proposed an introspective sensing framework that treats the response process itself as a sensing target in mental health tracking. By combining an explicit uncertain option '?' with decision latency measured from a smartwatch accelerometer, we showed that hesitation and uncertainty can be captured as behavioral indicators beyond conventional self-report. In an experiment using ten introspection-inducing questions, decision latency was longest for '?', and the three response types (immediate, internalized, and externalized) exhibited distinct patterns of symbol use and latency. These findings suggest that ambiguous responses and decision latency can enrich digital phenotyping by providing low-burden, stigma-sensitive information about how people arrive at their answers.

Several limitations should be noted. First, the three response types were defined heuristically, and their validity needs to be examined using more formal statistical methods. Second, in this study, we adopted a specific smartwatch interface with the symbolic structure '◯' (yes), '?' and '×' (no); individual characteristics such as social desirability, self-stigma, and perfectionism may affect how the uncertain option is used. Third, the experiment was conducted in an indoor environment; validation in everyday settings is required, including the effect of contextual factors, habituation through repeated use, and the temporal stability of response types. Addressing these issues and embedding the proposed framework into longitudinal mental health services will be an important step toward realizing practical introspective sensing in the wild.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP25K21258.

References

1 S. Shiffman, A. A. Stone, and M. R. Hufford: Annu. Rev. Clin. Psychol. **4** (2008) 1. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415
2 M. F. Armey, H. T. Schatten, N. Haradhvala, and I. W. Miller: Curr. Opin. Psychol. **4** (2015) 21. https://doi.org/10.1016/j.copsyc.2015.01.002
3 J.-P. Onnela and S. L. Rauch: Neuropsychopharmacology **41** (2016) 1691. https://doi.org/10.1038/npp.2016.7
4 J. Torous, J.-P. Onnela, and M. Keshavan: Transl. Psychiatry **7** (2017) e1053. https://doi.org/10.1038/tp.2017.25
5 A. Ikäheimonen, N. Luong, I. Baryshnikov, R. Darst, R. Heikkilä, J. Holmen, A. Martikkala, K. Riihimäki, O. Saleva, E. Isometsä, and T. Aledavood: J. Med. Internet Res. **26** (2024) e56874. https://doi.org/10.2196/56874
6 Y.-B. Shin, A. Y. Kim, S. Kim, M.-S. Shin, J. Choi, K. L. Lee, J. Lee, S. Byun, S. Kim, H.-J. Lee, and C.-H. Cho: BMJ Open **15** (2025) e096773. https://doi.org/10.1136/bmjopen-2024-096773
7 J. Haslbeck, A. J. Martínez, A. J. Roefs, E. I. Fried, L. H. J. M. Lemmens, E. Groot, and P. A. Edelsbrunner: Behav. Res. Methods **57** (2025) 217. https://doi.org/10.3758/s13428-025-02706-2
8 L. K. Kaye and C. R. Schweiger: Comput. Human Behav. **148** (2023) 107916. https://doi.org/10.1016/j.chb.2023.107916
9 B. S. Siepe, R. Tutunji, C. L. Rieble, R. K. K. Proppert, and E. I. Fried: J. Psychopathol. Clin. Sci. **134** (2025) 912. https://doi.org/10.1037/abn0001013
10 A. Pratap, D. C. Atkins, B. N. Renn, M. J. Tanana, S. D. Mooney, J. A. Anguera, and P. A. Areán: Depress. Anxiety **36** (2019) 72. https://doi.org/10.1002/da.22822
11 R. Büscher, T. Winkler, J. Mocellin, S. Homan, N. Josifovski, M. Ciharova, W. van Breda, S. Kwon, M. E. Larsen, J. Torous, and P. Cuijpers: npj Ment. Health Res. **3** (2024) 42. https://doi.org/10.1038/s44184-024-00089-4
12 V. M. Lawlor, C. A. Webb, T. V. Wiecki, M. J. Frank, M. Trivedi, D. A. Pizzagalli, and D. G. Dillon: Psychol. Med. **50** (2020) 1613. https://doi.org/10.1017/S0033291719001570
13 J. T. Nadler, R. Weston, and E. C. Voyles: J. Gen. Psychol. **142** (2015) 71. https://doi.org/10.1080/00221309.2014.994590
14 W. W. Gaver, J. Beaver, and S. Benford: Proc. SIGCHI Conf. Human Factors in Computing Systems (ACM, New York, 2003) p. 233. https://doi.org/10.1145/642611.642653
15 W. Gaver, P. Sengers, T. Kerridge, J. Kaye, and J. Bowers: Proc. SIGCHI Conf. Human Factors in Computing Systems (ACM, New York, 2007) p. 537. https://doi.org/10.1145/1240624.1240711
16 P. Sanches, K. Höök, E. Vaara, C. Weymann, M. Bylund, P. Ferreira, N. Peira, and M. Sjölinder: Proc. 8th ACM Conf. Designing Interactive Systems (ACM, New York, 2010) p. 47. https://doi.org/10.1145/1858171.1858182
17 D. Kahneman: Thinking, Fast and Slow (Farrar, Straus and Giroux, New York, 2011).
18 J. St. B. T. Evans and K. E. Stanovich: Perspect. Psychol. Sci. **8** (2013) 223. https://doi.org/10.1177/1745691612460685
19 K. Chung, J. Y. Park, D. Joung, and K. Jhung: JMIR mHealth uHealth **7** (2019) e14657. https://doi.org/10.2196/14657
20 O. P. John and S. Srivastava: Handbook of Personality: Theory and Research, L. A. Pervin and O. P. John, Eds. (Guilford Press, New York, 1999) 2nd ed., Chap. 2.
21 Android Developers: Sensors Overview (2025). https://developer.android.com/develop/sensors-and-location/sensors/sensors_overview (accessed September 6, 2025).
22 A. J. Lazard, B. A. Bamgbade, J. M. Sontag, and C. Brown: J. Health Commun. **21** (2016) 1260. https://doi.org/10.1080/10810730.2016.1245374