

Why don't you do what you said you would? Conversational strategies for agents to understand users' reasons in supporting behavior

Chen, P.Y.; van Riemsdijk, M. Birna; Heylen, Dirk K.J.; Jonker, C.M.; Tielman, M.L.

DOI

[10.1080/0144929X.2025.2483793](https://doi.org/10.1080/0144929X.2025.2483793)

Publication date

2025

Document Version

Final published version

Published in

Behaviour and Information Technology: an international journal on the human aspects of computing

Citation (APA)

Chen, P. Y., van Riemsdijk, M. B., Heylen, D. K. J., Jonker, C. M., & Tielman, M. L. (2025). Why don't you do what you said you would? Conversational strategies for agents to understand users' reasons in supporting behavior. *Behaviour and Information Technology: an international journal on the human aspects of computing*. <https://doi.org/10.1080/0144929X.2025.2483793>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Why don't you do what you said you would? Conversational strategies for agents to understand users' reasons in supporting behavior

Pei-Yu Chen, M. Birna van Riemsdijk, Dirk K. J. Heylen, Catholijn M. Jonker & Myrthe L. Tielman

To cite this article: Pei-Yu Chen, M. Birna van Riemsdijk, Dirk K. J. Heylen, Catholijn M. Jonker & Myrthe L. Tielman (17 Apr 2025): Why don't you do what you said you would? Conversational strategies for agents to understand users' reasons in supporting behavior, Behaviour & Information Technology, DOI: [10.1080/0144929X.2025.2483793](https://doi.org/10.1080/0144929X.2025.2483793)

To link to this article: <https://doi.org/10.1080/0144929X.2025.2483793>



© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 17 Apr 2025.



Submit your article to this journal [↗](#)



Article views: 170



View related articles [↗](#)



View Crossmark data [↗](#)

Why don't you do what you said you would? Conversational strategies for agents to understand users' reasons in supporting behavior

Pei-Yu Chen^a, M. Birna van Riemsdijk^b, Dirk K. J. Heylen^b, Catholijn M. Jonker^{a,c} and Myrthe L. Tielman^a

^aDepartment of Intelligent System, Interactive Intelligence Group, Delft University of Technology, Delft, The Netherlands; ^bDepartment of Computer Science, Human Media Interaction Group, University of Twente, Enschede, The Netherlands; ^cLeiden Institute of Advanced Computer Science (LIACS), Leiden University, Leiden, The Netherlands

ABSTRACT

Effective support from personal assistive technologies relies on accurate user models that capture user values, preferences, and context. Knowledge-based techniques model these relationships, enabling support agents to align their actions with user values. However, understanding values in a single context is insufficient due to the dynamic nature of behaviour. This study explores the use of dialogue strategies to update user models. Participants were randomly assigned to different strategies and they discussed one randomly chosen non-adherence situation with the agent. Then, their emotions, acquired information accuracy, completeness, and dialogue experience were rated. Our findings suggest that multiple-choice dialogues may limit response depth, reducing the perceived completeness of behaviour reasons. In contrast, open-ended questions allow more detailed input but require more time and effort, potentially worsening the dialogue experience. Through inductive coding, we identified key topics, such as individual challenges, priorities, tangible outcomes, and values, essential for constructing personalised user models. We also analyzed conversation paths to improve dialogue-based user model updates in support agents. Further research is needed to refine the relationship between dialogue strategies and self-conscious emotions, considering diverse backgrounds and health goals, while enhancing dialogue design.

ARTICLE HISTORY

Received 1 March 2024
Accepted 18 March 2025

KEYWORDS

Conversational agent;
behaviour change; values;
user modelling; alignment
dialogue

1. Introduction

Behavior support technology is designed to guide users' actions, for example, by means of support agents that help people eat healthier. Previous research shows that personalised approaches outperform generic 'one-size-fits-all' strategies in supporting behaviours (e.g. Krebs, Prochaska, and Rossi 2010; Lustria et al. 2013; Orji, Vasileva, and Mandryk 2014). Personalization encompasses using various forms of information about an individual, such as socio-demographic characteristics, and personality traits (Kaptein et al. 2015; Krebs, Prochaska, and Rossi 2010), to adjust the support. In this paper, we focus on acquiring information about the reasons for a user's behaviour in specific situations in a user model. Acquiring these reasons and the specific details of the situation is critical as it allows for a more accurate representation of the user and thereby provides support that aligns with the user's needs.

In contrast to data-oriented user models that use different learning techniques to predict or classify

users (Abri, Abri, and Çetin 2020), the user model we refer to employs knowledge-based techniques (Brachman 2004). These models explicitly capture the user's reasons and their relationships with desired behaviours. For example, a user model can describe the user's daily activities and which values are promoted or demoted by which activities (Tielman, Jonker, and van Riemsdijk 2018). If the user's behaviour reasons and specific situational details are modelled, the agent can select support actions that align with the user's needs when a similar situation occurs again.

However, there is a research gap in how to acquire the behaviour reasons and specific situational details for such a knowledge-based user model. Given the dynamic nature of user context and behaviours, we argue that interactive dialogues are promising because they facilitate the exploration of evolving user preferences and contextual nuances (Clark et al. 2019). Additionally, conversational agents offer benefits such as instant availability, a gentle learning curve, and platform independence

(Klopfenstein et al. 2017). In the same vein, Chen et al. (2023) proposed the concept of *human-agent alignment dialogues*, which are defined as dialogues with which the agent and user try to achieve or maintain alignment – support that aligns with the user’s values and preferences. Through these dialogues, the agent attempts to acquire what is significant to the user in their ever-changing daily contexts. As these contexts continually evolve, misalignments between the support provided by the agent and the user’s needs may occur. Alignment dialogues are proposed to address these misalignments, ensuring that the support remains relevant and effective.

Alignment dialogues include the user and the agent engaging in a question-answering conversation. The cycle continues until the misalignment no longer exists, allowing the agent to gain a better understanding of the current situation. For example, the agent might ask, ‘What factors influenced your decision to skip your workout today?’ or ‘How does the current situation affect your motivation to eat healthily?’ Different choices can be made about how these questions are phrased, ranging from open-ended to multiple-choice formats. The way these questions are structured can significantly impact the quality of the information gathered and the user’s emotional response.

In this study, we explore different dialogue strategies for alignment dialogues within the context of healthy lifestyle change, but the concept of alignment dialogues can be beneficial in other behaviour support domains too. These strategies vary in terms of the openness or closedness of the questions (detailed in Section 3.2). We focus on scenarios where the agent’s advice, although in line with the user’s goals as understood by the agent, is perceived as incorrect by the user. In such instances, the agent must capture the reasons behind the misalignment to identify if or how to revise the user model. This study explores how to structure a series of questions to acquire these reasons effectively.

One of the key behaviour-related aspects we aim to capture is users’ values. This is based on the premise that effective personalisation of technological support requires an understanding of what is important to users (Van Riemsdijk, Jonker, and Lesser 2015). Values are particularly useful in this regard, as they represent the criteria people use to make decisions and evaluate others and events (Friedman et al. 2013a; Van den Hoven, Vermaas, and Van de Poel 2015). Additionally, we are interested in how individuals prioritise different values across various situations. Therefore, we also consider contextual information about the situation and the social context to be useful.

In addition to eliciting useful information effectively, alignment dialogues should minimise the arousal of

negative emotions. As alignment dialogues often occur in situations where users fail to perform goal behaviours, negative emotions may arise (Fishbach and Zhang 2008; Ozkaramanli, Özcan, and Desmet 2017). Self-conscious emotions that are evoked by self-reflection and self-evaluation (Tangney, Stuewig, and Mashek. 2007) could be particularly relevant when the agent probes into reasons behind users’ non-adherence with goal behaviours (Chen et al. 2023).

In this study, we investigate the effectiveness of dialogues in eliciting the required information and which self-conscious emotions arise after these dialogues. Participants were randomly assigned to different dialogue strategies, discussed one randomly chosen non-adherence situation with the agent, and then rated their emotions, information accuracy, completeness, and dialogue experience. This work contributes to a deeper understanding of specific situational factors, in contrast to the general categories of barriers typically discussed in behaviour change literature, that prevent users from doing their goal behaviours. Specifically, we formulate the following research questions:

- RQ1: What are the effects of different dialogue strategies on the occurrence of self-conscious emotions?
- RQ2: What are the effects of different dialogue strategies on the dialogue experience, and on the subjective accuracy and completeness of the information acquired?
- RQ3: How effective are different dialogue strategies in eliciting the targeted information (social and contextual aspects of the situation, and underlying values)?

Additionally, we formulate two exploratory research questions:

- RQ4: Besides the targeted information, what additional topics do users mention?
- RQ5: What patterns emerge in users’ transitions between various topics when explaining their reasons for non-adherence behaviours?

The remainder of the paper begins with a discussion of Related Work (Section 2). Section 3 describes the design rationale of the dialogue strategies. We carried out a user experiment to investigate different dialogue strategies (Section 4). Subsequently, the results and discussion are presented in Sections 5 and 6. Insights derived from this study and their implications for the design of alignment dialogues are discussed (Section 7). We discuss future work (Section 8) and conclude the paper in Section 9.

2. Related work

In this section, we review existing literature relevant to our study, namely alignment dialogues (Section 2.1), and emotions that can arise in such interactions (Section 2.2). We clarify the unique perspectives of alignment dialogues, the possible topics these dialogues encompass, and the resulting potential impacts on the users.

2.1. Human-agent alignment dialogues and knowledge-based user modeling

Conversational agents have been developed to support a wide range of health-related activities, including behaviour change. These conversational agents largely focus on the actual support functionalities such as ‘instruction on how to perform a behaviour’ ‘social support’ and ‘problem-solving’ (Martinengo et al. 2022). These conversational agents were evaluated based on the effectiveness of the actual change of behaviours and the usability of the agents (Milne-Ives et al. 2020). However, in this study, the primary focus of the conversational agent is to gather information about the user by obtaining the reasons behind their non-adherence behaviours through *human-agent alignment dialogues*. This process aims to construct a more accurate and comprehensive user model. Such dialogues can be viewed as a necessary step before the agent provides actual support functionalities.

2.1.1. Defining alignment dialogues and their unique role

Human-agent alignment dialogues, as defined by Chen et al. (2023), are dialogues where the agent and user try to achieve or maintain alignment, which is defined as a situation where the provided support matches what the user needs or wants. This could be accomplished by the agent acquiring characteristics of users through the exchange in the dialogue. These dialogues are a form of task-oriented dialogue (Traum and Hinkelman 1992) but are different from traditional applications in conversational agent research in several aspects. First, common task-oriented dialogues focus on completing specific operational tasks, such as booking facilities (Griol et al. 2014). However, capturing high-level, situation-dependent concepts like values via a conversational agent, to the best of our knowledge, has not been extensively explored.

Meanwhile, in recent years there is a growing amount of research on conversational agents in the e-health domain. These agents typically focus on offering support through coaching, encouragement, and advice, facilitating various stages of behaviour change (Dingler et al. 2021; Hurmuz et al. 2020; Pereira and Díaz 2019;

Williams et al. 2012). In contrast, alignment dialogues have a different focus: rather than supporting the users in their health, alignment dialogues focus on ensuring the user model is correct. The underlying motivation is that only when the user model is a correct representation of the user, can the agent provide support that aligns with the user’s goals, norms, values, capabilities, and context.

To formally represent the connections between values, actions, and their relationships, Pasotti, Jonker, and van Riemsdijk (2017) have introduced a framework that models activities in hierarchies. Building upon this framework, Tielman, Jonker, and van Riemsdijk (2018) have demonstrated how values and contextual factors can be integrated. Values are linked to actions within a hierarchical structure. The model also specifies whether values are promoted or demoted by actions, and how a context influences the relationships.

2.1.2. Value elicitation

A primary focus of alignment dialogues is to elicit users’ values. Values are concepts that people use to evaluate people and events and make decisions (Friedman et al. 2013b; Van den Hoven, Vermaas, and Van de Poel 2015). Capturing these values enables technology to offer personalised support (Van Riemsdijk, Jonker, and Lesser 2015). Explicit elicitation of values is necessary (Berka et al. 2022) because inferring such high-level concepts from behavioural data is challenging (Armstrong and Mindermann 2018), especially considering that each of us has different values and that humans can behave irrationally (Butlin 2021). Moreover, values are abstract concepts that do not always carry the same meaning to all people. Therefore, incorporating values into a system should involve direct engagement with potential users during the design phase (can de Poel 2013).

Traditionally, values can be acquired explicitly through various questionnaires with predefined value lists, such as Schwartz (2012). However, directly asking individuals about their values can lead to incomplete or hypothetical responses that may not accurately represent real-life behaviours (Bostyn, Sevenhant, and Roets 2018). Moreover, these methods may acquire value preferences but are often not grounded in a context (Liscio et al. 2022; Pommeranz et al. 2011). Liscio et al. (2021) proposed a methodology to identify context-specific values by engaging human annotators in NLP tasks. While their work focuses on identifying values within a specific context, the term ‘context’ is used more broadly compared to how we consider the specific situation a user is in when making a behaviour-related decision.

In contrast, laddering interviews have been proven effective in eliciting deeper insights (Rugg et al. 2002;

Vanden Abeele, Hauters, and Zaman 2012). These interviews involve a series of ‘why’ questions to construct an attribute-consequence-value (ACV) chain (Miles and Rowe 2008). Laddering interviews are rooted in personality psychology and are applied to explore user needs and values (Berkovich et al. 2011; Rugg et al. 2002; Vanden Abeele, Hauters, and Zaman 2012). While traditionally used in requirements engineering for understanding consumer objectives, we conjecture that this method could be suitable for our context to guide users to describe the characteristics of the situation (‘attribute’) and its impact (‘consequence’), ultimately leading to a better understanding of their values. Inspired by the laddering interview technique, our first dialogue strategy employs a series of ‘why’ questions (Section 3.2.1).

2.1.3. Context and social awareness

While it is crucial to capture users’ values, we should also consider how individuals apply different values in varying situations. The prioritisation of values is heavily influenced by socio-cultural environments (Dignum 2017) and contexts (Hill and Lapsley 2009; Liscio et al. 2022). Therefore, besides capturing a user’s values, an agent should also identify the *contextual features* that influence behavioural decisions, such as non-adherence with the user’s stated goals. This aligns with the concept of context-awareness in system personalisation and tailoring.

Research in computer science uses terms such as situation-awareness (e.g. Endsley 1995) and context-awareness (e.g. Akman and Surav 1996) to describe approaches for enabling artificial agents to better understand their surrounding environment (Kola et al. 2020). op den Akker, Jones, and Hermens (2014) provides a comprehensive survey of various activity coaching systems that incorporate different forms of context-aware tailoring. op den Akker, Jones, and Hermens (2010) identify a set of optimal contextual features that enhance feedback compliance in physical activity systems.

A particular aspect of context-awareness is *social situation awareness*. This aspect specifically focuses on the user’s social context. Kola et al. (2022) defined a social situation as one involving multiple individuals. Social situation awareness emphasises factors like relationship quality and contact frequency between the user and others in the situation (Kola, Jonker, and van Riemsdijk 2019) and demonstrates how integrating social features into a decision tree can offer personalised and socially-aware behavioural support.

The contextual and social aspects of a situation inform our second and third dialogue strategies, in which we ask questions about these aspects, as well as the user’s values (Sections 3.2.2 and 3.2.3).

2.2. Self-conscious emotions

When it comes to behaviour change, a critical issue is user adherence with their behaviour goals. Research has demonstrated that when individuals are presented with choices that trigger both long-term goals and immediate desires, it often leads to internal conflicts (Fishbach and Zhang 2008). In such scenarios, choosing either option—adherence or non-adherence—can evoke both positive and negative emotions (Ozkaramanli, Özcan, and Desmet 2017).

Chen et al. (2023) highlight two aspects of behaviour support agents that may inadvertently provoke negative emotions in users: (1) The agent sometimes needs to confront the user’s non-adherence behaviours, and (2) The agent’s role in guiding the user towards their long-term goals, which may conflict with their immediate desires. The emotions commonly experienced by users in these situations include feelings of being judged and feelings of guilt, which are classified as self-conscious emotions (Tracy and Robins 2004).

Self-conscious emotions differ from basic emotions because they are evoked by self-reflection and self-evaluation (Tangney, Stuewig, and Mashek. 2007). Tracy and Robins (2004) proposed a process model explaining the emergence of self-conscious emotions, suggesting that individuals assess whether an event aligns with their goals and self-identity, which in turn influences the emotional outcome.

If the support provided by the agent is deemed appropriate from the agent’s perspective (i.e. aligns with the user’s long-term goals), but the user finds it unsuitable, it indicates a misalignment for the user. This incongruence could lead to negative self-conscious emotions. As a result, we consider self-conscious emotions to be a crucial measure in our study.

3. Dialogue design

In this study, we have chosen health-related behaviour changes, such as increasing physical activity or adopting a more nutritious diet, as our use case for studying different alignment dialogue strategies. In this section, we give an example scenario in this use case (Section 3.1) and describe the dialogues strategies used in the experiment (Section 3.2).

3.1. Use case scenario

Example Scenario 1. During Anna’s interaction with her behaviour support agent, she shares her goal of reducing sugar intake. The agent adjusts her online grocery options accordingly. However, Anna deviates from her routine and adds sugary items to her cart. This scenario

highlights the need for human-agent alignment dialogues as the agent's support action may not always align with the user due to incomplete information. For instance, Anna may have bought sweets for her birthday celebration, which the agent's model doesn't account for.

Example Scenario 2. Another user John, whose goal is to regularly attend the gym, skips several gym sessions. This scenario could initiate the agent to start an alignment dialogues to acquire John's reasons. For example, John might have skipped his workouts because he's facing a tight deadline at work, which takes precedence as it could impact his chances for a promotion.

The alignment dialogue aims to uncover such context, enabling the agent to update Anna's or John's user model and provide more relevant support.

3.2. Dialogue design choices

Following the identification of non-adherence scenarios, the agent engages in dialogue with the user to explore the underlying factors contributing to the deviation from their goals. Three dialogue strategies were used in the experiment, namely Exploratory Dialogues (Section 3.2.1), Focused Dialogues (Section 3.2.2), and Structured Choice Dialogues (Section 3.2.3).

3.2.1. Exploratory dialogues

Intuitively, to capture why the initial support was inadequate, the agent needs to probe into 'why' the support was wrong and elicit knowledge about what aspects of the current situation are significant to the user. In this dialogue variant, the agent aims to 'dig deeper' by posing a series of why-question, aiming to eventually uncover the user's values behind their decision in certain situations.

Exploratory dialogues involve the following questions. This setup was inspired by Rietz and Maedche (2023). Responses from the participants to these queries are in free-text form.

- (1) Q1: What particular factors related to the situation influence your choice?
- (2) Q2: Why is it important to you?
- (3) Q3: What feeling does it give you?
- (4) Q4: Why is that?

3.2.2. Focused dialogues

Focused Dialogues employ a dialogue strategy that involves using open questions but with a targeted information-seeking approach. This selection of targeted information was based on our assumptions, drawn

from existing literature about the aspects influencing user behaviour, namely: social aspects (Kola et al. 2022), contextual aspects (op den Akker, Jones, and Hermens 2014), and personal values (Schwartz 2012; Van Riemsdijk, Jonker, and Lesser 2015).

The agent initiates the dialogue with the user by outlining that it will explore three aspects of the scenario, accompanied by the definition from Kola et al. (2022), op den Akker, Jones, and Hermens (2014), and Friedman et al. (2013a) for social aspects, contextual aspects, and personal values respectively. Responses from the participants to these queries are in free-text form. This intentional structure progresses from tangible details to broader considerations, encouraging introspection.

- (1) Social aspects: These encompass the elements within a situation that are related to social interactions, relationships, and social dynamics.
Q1: What social aspects in this situation contribute to your choice?
- (2) Contextual aspects: These can be any information within this situation that can influence, shape, or affect your choice.
Q2: What contextual aspects in this situation contribute to your choice?
- (3) Personal values: Personal values are what a person considers important in life.
Q3: What personal values contribute to your choice in this situation?

3.2.3. Structured choice dialogues

In addition to two open-ended strategies, we employed a structured approach by presenting participants with lists of predefined social aspects, contextual aspects, and values as multiple-choice. The questions in Structured Choice are identical to those posed in the Focused Dialogue. The difference lies in Structured Choice Dialogues offer participants predetermined options to select from. This method provides participants with predefined options, allowing them to choose the elements most relevant to their decision. Below we describe the lists used.

- (1) *Social aspects*: This list was adapted from Kola, Jonker, and van Riemsdijk (2019), who formalise elements of social situations, including the role a person plays, how often they are in touch with others, the nature of relationships, depth of acquaintance, quality of the relationship, and how formal or informal they are with those around them.
- (2) *Contextual aspects*: For contextual aspect, we used the lists from op den Akker, Jones, and Hermens (2010, 2014), which capture elements such as the weather conditions, the proximity to places where

goals can be pursued, timing-related details (like day of the week, specific occasions), and past behaviours concerning the goal.

- (3) *Personal values*: We used the List of Values (LOV) for presenting values (Kahle 1983). These values from the LOV are known to be related to consumer activities and healthy lifestyle choices (Divine and Lepisto 2005; Homer and Kahle 1988). The LOV includes the importance of interpersonal relations, personal factors (e.g. self-respect, self-fulfillment), and personal enjoyment (e.g. fun, excitement) in value fulfillment.

3.3. Hypotheses

Based on the reviewed literature and the dialogue strategy design, we articulated the following hypotheses, corresponding to RQ1 to RQ3.

- H1: Participants in Exploratory Dialogues will experience greater self-conscious emotions compared to Focused and Structured Choice Dialogues, as open-ended questions may be perceived as more judgmental (Chen et al. 2023).
- H2a: Exploratory Dialogues will result in a lower dialogue experience than Focused Dialogues or Structured Choice Dialogues. Open-ended questions may require more time and effort to answer (Ben-Nun 2008; Chen 2017), potentially inducing cognitive burden and leading to skipped or low-quality responses (Ben-Nun 2008; Chen 2017; Reja et al. 2003).
- H2b: Structured Choice Dialogues will lead to lower subjective accuracy and completeness compared to Exploratory and Focused Dialogues. This is because the closed nature of Structured Choice could limit users' ability to express themselves in a manner that feels accurate to them (Chen 2017) and that users may find the use of values unrelatable or unnatural (Bostyn, Sevenhant, and Roets 2018).
- H3: Structured Choice Dialogues will be the most effective in eliciting users' values, social, and contextual aspects, followed by Focused Dialogues, with Exploratory Dialogues being the least effective.

4. Method

A user experiment was conducted to investigate how effective different dialogue strategies are in eliciting information related to contextual and social aspects of

a situation, and underlying values (RQ1). In addition, an inductive coding method was used to see if additional topics came up in the Exploratory Dialogues and Focused Dialogues (RQ2). Validated questionnaires were used to measure self-conscious emotions and dialogue experience (RQ3). Finally, participants were asked to rate the accuracy and completeness of the summary of their provided input (RQ4).

4.1. Participants

Participants were recruited from the crowdsourcing platform Prolific¹, where they received monetary compensation according to platform policies. Approval for the study was obtained from [Anonymous]. A total of 234 participants were included in the analysis. 82.91% of individuals were below 35 years old, 14.10% were between 35 and 54 years old, and 2.99% were 55 years old or older. Educational backgrounds varied, with participants holding different levels of education (high school 30.34%, Bachelor 55.98%, Master 12.39%, PhD 0.85%). Additionally, 46.15% reported degrees or professional experience in STEM fields, while computer programming skills varied (no experience 32.48%, novices 37.18%, competent 27.77%, experts 2.56%).

4.2. Procedure

We used two platforms for the experiment: Qualtrics² for the questionnaires and Landbot³ for testing the dialogues. Landbot is a rule-based platform that allows us to prototype and test different dialogue strategies without having to build a conversational agent from scratch. It follows the dialogue flows we created, as detailed in Section 3.2. The experiment had two main parts:

Participants first completed a Qualtrics survey via Prolific, providing background information, computer skills, and emotion traits (see Section 4.3.1). Subsequently, they engaged with a Landbot conversational agent. During the chat, they shared one health-related goal and three *non-adherence situations* where they thought it might be difficult to stick to their goals. They also rated their motivation level for the goal and the perceived realism of the non-adherence situations.

Participants were randomly assigned to different dialogue strategy conditions. While they initially provided three non-adherence situations, only one was discussed with the agent during the second part, chosen at random to simulate real-life scenarios. The questions they were posed during the human-agent dialogue in Part 2 were based on the assigned dialogue strategy. After the dialogue, participants returned to Qualtrics to rate

their self-conscious emotions, accuracy and completeness of the acquired information, and their dialogue experience. We detail these measures in the next section.

4.3. Measures

In this section, we detail the instruments employed to evaluate the study. These measures include control variables (Section 4.3.1), self-conscious emotions measures (Section 4.3.2), the Dialogue Experience Questionnaire (DEQ) (Section 4.3.3), and assessments of accuracy and completeness (Section 4.3.4).

4.3.1. Control variables

Demographics. Participants provided essential demographic information, including age, gender, and highest level of education. They also rated their technological proficiency on a 4-point scale from 'No experience' to 'Expert,' and were asked about their academic or professional experience in STEM-related disciplines.

Self-conscious emotion proneness. To establish a comparative baseline for emotional responses, we assessed participants' inclination towards self-conscious emotions using the Test of Self-Conscious Affect-2 (TOSCA-2) (Tangney et al. 2000). Participants rated their likelihood of reacting to hypothetical scenarios on a 5-point scale, focussing on Shame, Guilt, Authentic Pride, and Hubristic Pride-Proneness. This measurement helped compare proneness with actual emotional responses post-interaction.

Motivated & Realistic Assessment. Additionally, after each participant shared their individual healthy lifestyle goal and identified three challenging situations that could hinder adherence, the conversational agent prompted them to evaluate, on a scale ranging from 1 (not at all) to 5 (extremely), both their level of motivation to accomplish their goals and the perceived realism of encountering the obstructive situations.

4.3.2. Self-conscious emotions measures

Given our focus on behaviour change for a healthy lifestyle, we employed the Body-related Self-Conscious Emotions Fitness instrument (BSE-FIT) (Castonguay et al. 2016). Participants rated 16 statements on a 5-point scale from 'strongly disagree' to 'strongly agree.' Adaptations were made to inquire about emotions triggered by the conversational agent, e.g. changing words to assess emotion 'at the moment' instead of general.

4.3.3. Dialogue experience questionnaire (DEQ)

We assessed users' alignment dialogue experiences using the Dialogue Experience Questionnaire (DEQ) (ter Heijden and Brinkman 2011). We mainly focussed on the 'Interaction: discussion satisfaction' section. In addition, we included two items from the 'Interaction: reality' section: *The conversation with the agent felt natural* and *I had to adjust my way of interacting to communicate with the agent*. Participants answered on a 7-point Likert scale, ranging from 'Strongly disagree' to 'Strongly agree.'

4.3.4. Accuracy and completeness

To assess how well the dialogue acquired the reasons for non-adherence, the agent compiled a recap at the end of each interaction, template-filling the information provided by the participants. Participants then rated on a scale from 1 (not at all) to 5 (extremely) the accuracy and completeness of this recap in acquiring their reasons for non-adherence.

4.4. Text responses analysis method

Exploratory Dialogues and Focused Dialogues yielded free-text responses from participants, which we analyzed using a combination of deductive and inductive analysis (Fereday and Muir-Cochrane 2006) (Sections 4.4.1 and 4.4.2), and summarise the coding process (Section 4.4.3).

4.4.1. Deductive coding

The deductive coding approach involves predefined coding schemes developed by previous research or theory to categorise data (Crabtree and Miller 1992). However, existing behaviour change research primarily focuses on facilitators and barriers related to individuals' characteristics, views, or beliefs (Deslippe et al. 2023; Sullivan and Lachman 2017), often overlooking the reasons behind behaviours in specific contexts, such as the non-adherence scenarios central to our study. Consequently, we developed our own coding scheme. First and foremost, we examined whether participants introduced *novel information* in their answers—defined as any new details or insights that the user has not previously mentioned at any point during the dialogue. This definition does not limit the novelty to the immediately preceding response; rather, it encompasses all parts of the dialogue, regardless of whether it pertains to social, contextual, or personal value aspects. These three aspects were only annotated if an utterance was first coded as novel information. Our deductive coding scheme included four categories: novel information, social aspects, contextual aspects, and personal values.

We annotated the utterances with these predefined codes as either being mentioned or not.

4.4.2. Inductive coding

If a response coded as having novel information did not fit under the codes of social aspects, contextual aspects, or personal values, we used inductive coding. In this process, we analyzed the data to identify and establish new codes based on key points within the responses. We followed the inductive coding stages from Thomas (2003).

4.4.3. Coding process

In summary, our text analysis method integrates both deductive and inductive coding approaches. We follow the following coding procedure. Additionally, an independent coder evaluated one third of the data. The independent coder was given the definitions of novel information, social information, contextual information, and personal values. The independent coder was also given examples of utterances that should be coded as such. In instances where coders disagreed on the coding of an utterance, they engaged in further discussion to arrive at a consensus.

- (1) Each utterance is assessed to determine if it contains ‘novel information’ in comparison to what participants had said.
- (2) If no novel information is identified, the coding process concludes for that utterance.
- (3) If novel information is present, proceed to annotate whether it relates to predefined codes, such as social, contextual, or personal values (deductive coding).
- (4) Additionally, we annotate any emerging topics the utterance contain, applying inductive coding.

To summarise, across all dialogue strategies, we measure self-conscious emotions, DEQ, accuracy, and

completeness. Additionally, we coded the responses in both the Exploratory Dialogues and Focused Dialogues using deductive coding. For the Exploratory Dialogues, we further employed inductive coding, as we were particularly interested in identifying and merging topics that did not fall under social, contextual, or personal values. Figure 1 shows an overview of the study design and data analysis.

5. Results

A thorough assessment of the control variables—namely, gender, age, education, and emotion proneness—indicated no significant variations across the different dialogue strategy groups.

5.1. Self-Conscious emotions analysis

To answer RQ1, ‘What are the effects of different dialogue strategies on the occurrence of self-conscious emotions?’, we first calculated the averages and standard deviations of each emotion in different dialogues, which are shown in Table 1 and illustrated by Figure 2. These demonstrate the participants’ self-conscious emotions in response to various dialogue strategies after engaging in alignment dialogues. On average, across all dialogue conditions, participants reported the highest emotional experience of Guilt (mean = 2.90), followed closely by Authentic Pride (mean = 2.79), then Shame (mean = 2.5), with the least being Hubristic Pride (mean = 2.48).

Furthermore, we applied a linear regression model to examine variables’ influence, including the participants’ self-conscious emotion proneness, their motivation levels to engage in the goal behaviours, and how realistic the challenging situations (detailed in Section 4.3.1). Our predictor selection employed a stepwise procedure that merged both forward and backward selections, ensuring the identification of the optimal model.

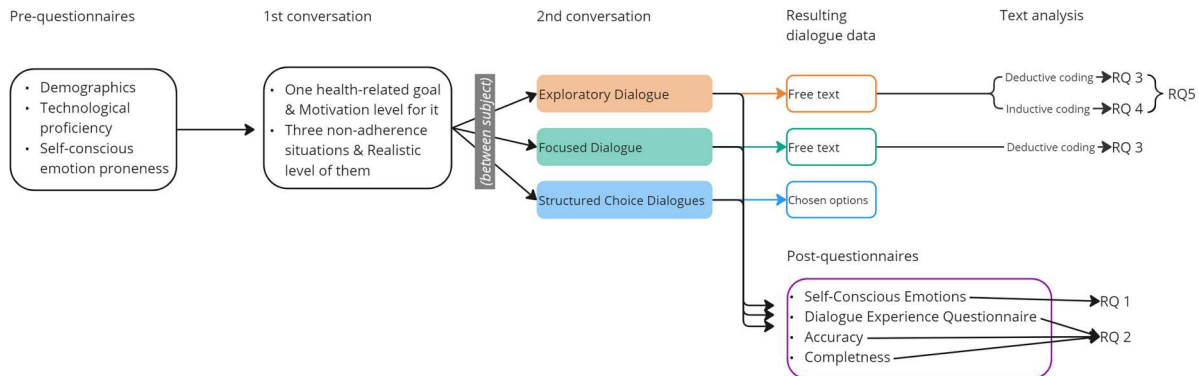


Figure 1. Overview of the study design and data analysis.

Table 1. Mean (SD) of self-conscious emotions reported by participants across different dialogue strategies, measured on a scale of 1 to 5 (1 indicating 'strongly disagree' and 5 indicating 'strongly agree').

	Shame	Guilt	Authentic Pride	Hubristic Pride
Exploratory	2.66 (0.10)	3.02 (0.12)	2.57 (0.10)	2.36 (0.11)
Focused	2.45 (0.11)	2.95 (0.12)	2.89 (0.11)	2.48 (0.09)
Structured Choice	2.42 (0.11)	2.74 (0.12)	2.90 (0.10)	2.60 (0.10)
Overall	2.51 (0.93)	2.90 (1.05)	2.79 (0.92)	2.48 (0.90)

Shame. The regression analysis indicated that participants' shame proneness was a significant predictor of the level of shame they reported experiencing during the alignment dialogues ($\beta = 0.46, p < .001$). Other factors did not emerge as significant predictors in the final model.

Guilt. Similarly, the regression analysis showed that participants' shame proneness was a significant predictor of the guilt experienced during the alignment dialogues ($\beta = 0.43, p < .001$).

Authentic pride. Our analysis showed the dialogue strategies and motivation level to be significant predictors of feelings of authentic pride. Specifically, participant motivation to achieve their goals was positively associated with feelings of authentic pride ($\beta = 0.27, p < .001$). Participants in the Exploratory Dialogues group felt less authentic pride than those in the Focused Dialogues group ($\beta = -0.28, p < .05$). However, there was no clear difference in feelings of authentic pride between the Focused Dialogues and Structured Choice Dialogues groups, and we did not find any interaction between the dialogue strategies and motivation levels.

Hubristic pride. For hubristic pride, the regression analysis indicated participants' motivation to achieve their goals as a significant predictor of hubristic pride ($\beta = 0.23, p < .001$).

5.2. Dialogue experience questionnaire (DEQ), accuracy and completeness

RQ2 addresses the effects of different dialogue strategies on the dialogue experience, as well as on the subjective accuracy and completeness of the information acquired. The results are descriptively summarised in Table 2, providing insights into how each strategy influences these aspects of user interaction.

We applied the same linear regression method used for emotions. In our analysis, we included dialogue strategy, motivation levels toward achieving the goals, and perceptions about the likelihood of encountering non-adherence scenarios. We also considered background factors like education level, computer programming skills, and possession of a STEM degree, believing that technological expertise might influence their assessment of the dialogue.

DEQ. Our regression analysis revealed that participants' perception of how realistic the scenarios were, was negatively associated with their DEQ scores ($\beta = -0.17, p < .05$). Additionally, participants' motivation levels were positively associated with their DEQ scores ($\beta = 0.33, p < .001$). The dialogue strategy used had an impact too. Specifically, participants in the Exploratory Dialogues condition reported significantly lower DEQ scores than those in the Focused Dialogues condition ($\beta = -0.41, p < .05$). No significant difference in DEQ scores was observed between the Focused

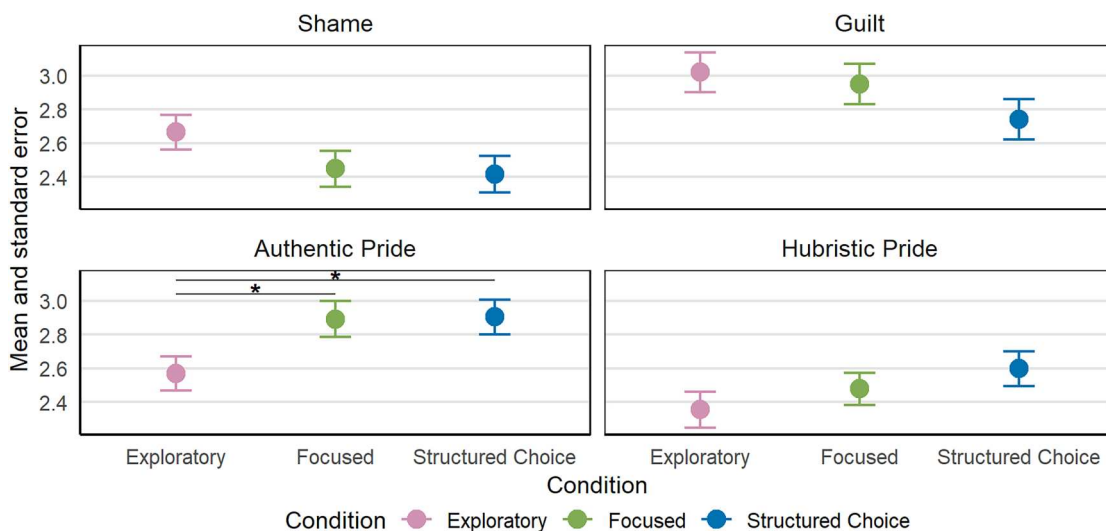


Figure 2. Plot of emotions by dialogue strategy with error bars.

Table 2. Descriptive statistics of dialogue experience questionnaire (DEQ), measured on a scale from 1 (Strongly Disagree) to 7 (Strongly Agree) with statements regarding satisfaction and realism, Accuracy, rated on a scale 1 (Not at all accurate) – 5 (Extremely accurate), and Completeness, rated on a scale 1 (Not at all complete) – 5 (Extremely complete).

	DEQ	Accuracy	Completeness
Exploratory Dialogues	4.13 (0.93)	4.03 (1.13)	4.12 (0.98)
Focused Dialogues	4.37 (0.87)	4.06 (0.86)	4.17 (0.78)
Structured Choice Dialogues	4.41 (0.86)	3.96 (0.82)	3.78 (0.87)
Overall	4.30 (0.89)	4.02 (0.94)	4.02 (0.89)

Dialogues and Structured Choice Dialogues groups. Interestingly, education levels themselves did not have a main effect. However, after aggregating these into two main categories: ‘High school’ and ‘Above Bachelor’ (which combines Bachelor, Master, and PhD levels), we found significant interactions between education level and dialogue strategies. The interaction effect is demonstrated in Figure 3.

Table 3 shows the DEQ scores in different dialogue strategies with different educational levels. The interaction between the Structured Choice Dialogues and High school was significant ($\beta = 0.67, p < .05$). Similarly, the interaction between Exploratory Dialogues and High school also showed significance ($\beta = 0.60, p < .05$).

Accuracy. We observed a significant correlation between participants’ accuracy ratings and their perceived realism of non-adherence scenarios ($r = 0.19, p < .05$).

Completeness. Regarding the completeness of the reasons, we found a significant correlation with participants’ motivation levels ($r = 0.18, p < .01$) and the dialogue strategies. Specifically, participants in the

Table 3. Dialogue experience questionnaire (DEQ) scores across different dialogue strategies by educational level.

	High school		Bachelor and above	
	n	Mean (SD)	n	Mean (SD)
Exploratory Dialogues	17	4.43 (0.72)	58	4.02 (0.97)
Focused Dialogues	24	4.25 (0.86)	54	4.42 (0.88)
Structured Choice Dialogues	30	4.55 (0.67)	50	4.33 (0.95)

Structured Choice Dialogues condition reported significantly less complete reasons compared to those in the Focused Dialogues condition ($r = 0.18, p < .01$). No significant difference in completeness level was observed between the Focused Dialogues and Exploratory Dialogues groups, and there was no interaction effect between dialogue strategies and motivation levels. Table 4 presents the summary of all the regression models.

5.3. Text responses analysis

In this section, we present the results of text analysis. Deductive coding reports the occurrences of predefined codes, answering RQ3 (Section 5.3.1) and inductive coding reveals emergent codes, answering RQ4 (Section 5.3.2).

5.3.1. Deductive coding

To address RQ3, ‘How effective are different dialogue strategies in eliciting targeted information?’, we calculate the percentage of responses that yielded novel information. Then, within this set of responses containing novel information, we further calculate the percentages that pertain to predefined categories of interest: social factors, contextual factors, and personal values. Table 5 provides a detailed breakdown of these

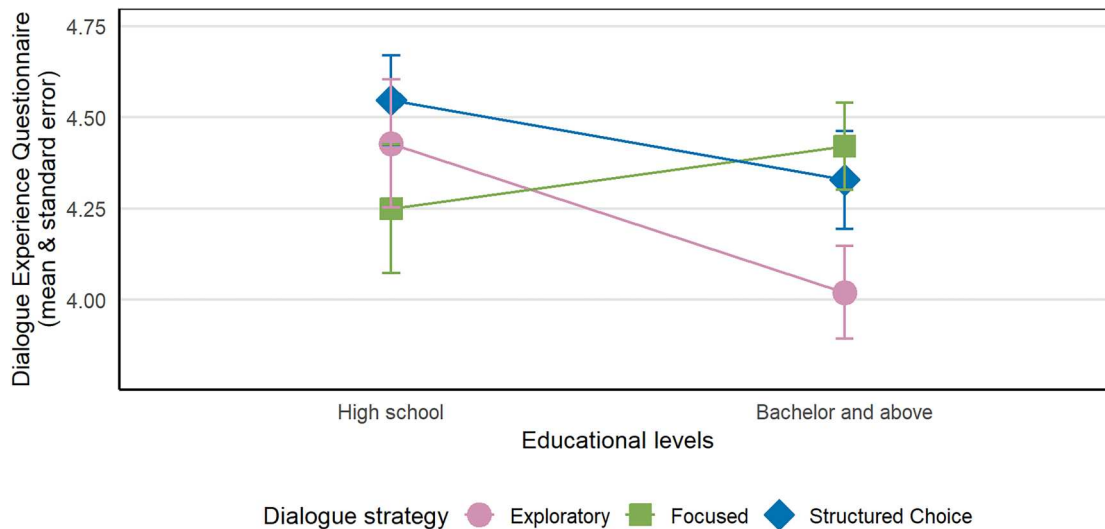


Figure 3. Interaction plot of DEQ by aggregated education and condition (Mean with error bars).

Table 4. Final regression models for assessing the effects of dialogue strategies on self-conscious emotions, accuracy, completeness, and DEQ.

Measures	Variables (Predictors)	β	SE	p value
Shame	Shame proneness	0.43	0.10	< .001***
Guilt	Shame proneness	0.46	0.11	< .001***
Authentic pride	Dialogue strategy: Structured Choice	−0.02	0.14	0.91
	Dialogue strategy: Exploratory	−0.28	0.14	< .05*
	Motivation level	0.27	0.06	< .001***
Hubristic pride	Motivation	0.23	0.06	< .001***
Accuracy	Realistic level	0.18	0.07	< .05*
Completeness	Dialogue strategy: Structured Choice	−0.41	0.14	< .01**
	Dialogue strategy: Exploratory	−0.02	0.14	< 0.89
	Motivation level	0.18	0.06	< .01***
DEQ	Dialogue strategy: Structured Choice	−0.25	0.16	< 0.13
	Dialogue strategy: Exploratory	−0.4	0.16	< .05*
	Education: High school	−0.20	0.20	0.32
	Realistic level	−0.17	0.07	< .05*
	Motivation level	0.33	0.06	< .01***
	Dialogue strategy: Structured Choice* Education: High school	0.67	0.28	< .05*
	Dialogue strategy: Exploratory* Education: High school	0.60	0.30	< .05*

Note: The table only displays predictors with significant effects. For categorical variables (e.g. dialogue strategy), all levels are shown for clarity and to indicate the reference level used in comparisons.

percentages, showcasing the effectiveness of each dialogue strategy in eliciting the predefined information.

For Exploratory Dialogues, the proportion of responses containing novel information varied across the dialogue questions (Q1 to Q4), with Q1 having the highest at 83%, followed by Q3 at 79%, Q2 at 62%, and Q4 at 27%. Social and contextual factors were most frequently mentioned in Q1, though less so in subsequent questions. Interestingly, while Q1 showed little mention of personal values, this topic was most predominant at 29% of novel information in Q4, with 23% of all novel information responses in Q2.

For Focused Dialogues, the novel information was predominantly in the answer to Q1 at 82%, out of which 48% provided information on this aspect. This was followed by Q2 at 56% (out of which 27% provided contextual information) and Q3 at 51% (out of which 30% provided value information).

5.3.2. Inductive coding

To address RQ4 –‘Besides the targeted information, what additional topics do users mention?’ –we

Table 5. Deductive coding in exploratory dialogues and focused dialogues.

	Exploratory Dialogues				Focused Dialogues		
	Q1	Q2	Q3	Q4	Q1	Q2	Q3
Novel information	0.83	0.62	0.79	0.27	0.82	0.56	0.51
Social aspects	0.06	0	0	0.10	0.48	–	–
Contextual aspects	0.13	0.02	0	0	–	0.27	–
Personal values	0.03	0.23	0.03	0.29	–	–	0.30

Note: The first row indicates the percentage of responses containing novel information, calculated by dividing the number of responses with novel information by the total number of responses. The following rows indicate the percentage of these novel information responses that further discuss social, contextual, and personal value aspects, respectively, calculated by dividing the number of responses addressing each aspect by the total number of responses containing novel information. See Section 3 for what Q1–Q4 refers to in each dialogue condition.

employed inductive coding in the Exploratory Dialogues. This approach was used to identify topics in responses that did not fall under the targeted categories of social aspects, contextual aspects, or personal values. Below, we present a list of these emergent topics with explanations and illustrative examples from the responses.

- (1) *Individual Challenges* encompass intrinsic factors that are inherent to the individual and are difficult to ask of them to change.
- (2) *Environmental and Societal Challenges* refer to broader, external factors that shape user behaviour. These challenges are generally stable and persistent, rather than tied to the user’s immediate circumstances.
- (3) *Prioritization and Preferences* underscore the influence of personal tastes, likes, dislikes, preferences and attitudes.
- (4) *Norms and Obligations* cover actions and choices driven by societal/personal norms or obligations.
- (5) *Beliefs* encompass a wide range of beliefs, such as cultural beliefs and factual beliefs.
- (6) *Tangible Outcomes* encompass the practical, observable consequences of behavioural decisions.
- (7) *Emotional Outcomes* relate to affective responses resulting from (lack of) behaviour.

5.4. Patterns in user explanations

To answer RQ5 ‘What patterns emerge in users’ transitions between different topics when they explain their reasons for non-adherence behaviours?’, we analyzed the conversation paths in the Exploratory Dialogues which start with a user’s non-adherence scenario

and then four questions from the agent. The primary objective was to identify patterns in how users transitioned between various topics of explanations. Our goal was to understand these transition patterns better, thereby aiding in the selection of effective questions for eliciting informative responses. Below is the series of questions the agent asked after the participant provided the non-adherence scenario:

(0) *User's Non-adherence Scenario*

- (1) Q1: What particular factors related to the situation influence your choice?
- (2) Q2: Why is it important to you?
- (3) Q3: What feeling does it give you?
- (4) Q4: Why is that?

For the non-adherence scenario and each response to a question, we formed four pairs of consecutive responses which we refer to as 'Duos.' We adopted two distinct perspectives for analyzing the Duos: Preceding Path Analysis and Following Path Analysis. The former examines the percentage of topics *preceding* a certain topic, while the latter analyzes the percentage of topics *following* a certain topic. In other words, Preceding Path Analysis can reveal, for instance, that topic x is most often preceded by topic y. If the agent wants information about topic x, they could lead with topic y. Conversely, Following Path Analysis can show that topic y often follows topic x. Asking about topics y after topic x could make the conversation feel more natural to the users.

For the Preceding Path Analysis, we calculated the percentage of instances where topic x was led by each specific topic out of all topics that led topic x. For the Following Path Analysis, we calculated the percentage of instances where each specific topic followed topic x out of all topics that followed topic x. We applied these two approaches differently depending on the nature of the dialogues. Below are the results of each Duo.

- (1) *Duo (Scenario, Q1)*: For the Duo that goes from the start to Q1, we exclusively employed the Following Path Analysis. This choice was made because our interest lay in understanding the explanations of why these scenarios led to their non-adherence behaviours. Table 6 shows the top explanations that were at least 10% in the analysis. Below is a list of summary.

- Detailed Challenges: Participants often provide detailed explanations of challenges following scenarios involving Environmental and Societal Challenges or Individual Challenges.

Table 6. Following path analysis of duo (Scenario, Q1).

Scenario	Q1	Percentage
Env. Soc.	Env. Soc.	0.27
	Pref.	0.13
	Indiv. Contxt.	0.09
Indiv.	Indiv.	0.33
	Env. Soc.	0.33
	Pref.	0.22
Motiv.	Pref.	0.16
Pref.	Pref.	0.4

Note: Only the top few topics with a cutoff point around 0.1 are shown (Abbreviations used: 'Indiv.' for Individual Challenges, 'Contxt.' for Contextual Information, 'Env. Soc.' for Environmental and Societal Challenges, 'Pref.' for Prioritization and Preferences, 'Norm. Obl.' for Norms and Obligations, 'Tang.' for Tangible Outcomes.)

- Prioritization and Preferences consistently emerge as prominent topics across various scenarios, revealing participants' considerations in prioritising or preferring elements over behaviour goals.
 - Contextual Information is in some cases provided, showing situational factors influencing behaviour decisions.
- (2) *Duo (Q1, Q2)*: we utilised both the Following Path Analysis and the Preceding Path Analysis. Using both analyses allowed us to understand the topics both preceding and following the participants' responses.
 - (a) Following Path Analysis (Table 7)
 - After discussions on Environmental and Societal Challenges or Individual Challenges, users often move to topics related to Tangible Outcomes or Prioritization and Preferences.
 - Following discussions on Prioritization and Preferences, users frequently transition to Tangible Outcomes, indicating an intertwined nature of these topics.
 - When users discuss Tangible Outcomes, Norms and Obligations, Social Information, or Contextual Information, the subsequent topic tends to be Values, suggesting a likely progression in the conversation.
 - (b) Preceding Path Analysis (Table 8)

Table 7. Following path analysis of duo (Q1, Q2).

Q1	Q2	Percentage
Env. Soc.	Tang.	0.23
	Pref.	0.13
Indiv.	Pref.	0.15
	Tang.	0.15
Pref.	Tang.	0.27
	Pref.	0.19
Tang.	Value	0.29
Norm. Obl.	Value	0.40
Social	Value	0.43
Contextual	Value	0.33

Note: See Table 6 for the abbreviations used.

Table 8. Preceding path analysis of duo (Q1, Q2).

Q2	Q1	Percentage
Env. Soc.	Env. Soc.	0.33
Pref.	Pref.	0.33
	Pref.	0.27
Tang.	Env. Soc.	0.23
	Pref.	0.23
Beliefs	Env. Soc.	0.25
Norm. Obl.	Env. Soc.	0.43
Value	Social	0.23
	Contextual	0.16
	Env. Soc.	0.16

Note: See Table 6 for the abbreviations used.

- Some topics in Q2 are most frequently preceded by the same topics in Q1, such as Prioritization and Preferences, as well as Environmental and Societal Challenges.
 - Tangible outcomes are preceded most by Environmental and Societal Challenges & Prioritization and Preferences.
 - Norms and Obligations, as well as Beliefs are predominantly preceded by Environmental and Societal Challenges.
 - Values are mostly preceded by Environmental and Societal Challenges, Context, or Social factors.
 - Above observations imply that initiating the dialogue with questions related to Environmental and Societal Challenges or Prioritization and Preferences could effectively lead to discussions about Tangible outcomes, Norms and Obligations, Beliefs, or Values.
- (3) *Duo* (Q2, Q3): We skipped this due to Q3's focus on emotions, making preceding or following analysis unnecessary.
- (4) *Duo* (Q3, Q4): We employed Following Analysis, as Q3 focussed on emotions, we were interested in what topics followed the emotional responses.
- After discussing emotions, users provided 'Value Information' more than other topics. This trend suggests that probing emotional consequences could lead users to articulate their values.

6. Discussion

This study explored how dialogue strategies affect self-conscious emotions, user satisfaction – comprising accuracy, completeness, and dialogue experience – and the acquisition of the user reasons behind non-adherence in specific scenarios. In the following sections, we discuss our findings related to self-conscious emotions (Section 6.1) and user satisfaction (Section 6.2). Subsequently, we summarise the qualitative information regarding reasons for non-adherence behaviours (Section 6.3). Finally, we highlight other noteworthy

observations that emerged during our study (Sections 6.4 & 6.5).

6.1. Emotions in different dialogue strategies

Our analysis reveals that emotions of shame, guilt, and hubristic pride remain largely consistent across the different dialogue strategy conditions (see Figure 2). This suggests that the dialogue strategies did not have a significant impact on the manifestation of these self-conscious emotions.

However, a notable observation is that participants in the Exploratory Dialogues condition reported lower levels of authentic pride in comparison to those in the Focused Dialogues and Structured Choice Dialogues conditions. Authentic pride is often associated with feelings of accomplishment and the motivation to pursue goal-oriented actions (Carver, Sinclair, and Johnson 2010; Castonguay et al. 2015, 2016). The straightforwardness inherent in the Exploratory Dialogues might give participants the impression that the agent is subtly challenging the value or importance of their goals. This finding may be attributed to the fully open-ended nature of the why-questions in the Exploratory Dialogues, which may be perceived as seeking justification. The ambiguous nature of 'why' questions may lead to unproductive or even hostile responses. Dewdney and Michell (1997). Conversely, the Focused Dialogues and Structured Choice Dialogues strategies, which employ more specific and targeted questions, could be perceived as being less direct and intrusive.

To summarise the implications with respect to self-conscious emotions: individual shame proneness is a clear predictor of shame and guilt. Beyond that, motivations seem intertwined with authentic pride and hubristic pride. How the open questions are framed in the Exploratory Dialogues strategy seems to negatively impact authentic pride.

6.2. DEQ, accuracy, and completeness in different dialogue strategies

Regarding user satisfaction, our observations reveal that there were not any significant differences in the accuracy of the reasons for non-adherence. However, when it comes to the completeness of the reasons, participants in the Structured Choice Dialogues tended to perceive the summary as significantly less complete compared to their counterparts in the Exploratory Dialogues or Focused Dialogues. This result aligns with our expectations since Structured Choice Dialogues participants were presented with predefined multiple-choice options

that covered social, and contextual aspects of the situation, and values. The restrictive nature of these options may have hindered participants from providing more expansive responses, given the inherent limitations of the setup (Chen 2017).

Furthermore, we identified a significant decrease in the dialogue experience score (DEQ) for the Exploratory Dialogues compared to the Focused Dialogues and Structured Choice Dialogues conditions. This trend mirrors the patterns observed with authentic pride. Further examination revealed a strong correlation between authentic pride and DEQ ($\beta = 0.4, p < .001$), reinforcing the interplay between the dialogue strategy, and resultant pride emotions and DEQ. This might be because Exploratory Dialogues demand additional time and cognitive effort (Ben-Nun 2008; Chen 2017), contributing to a worse dialogue experience compared to other conditions.

To summarise the implications on user satisfaction: the realistic level of non-adherence scenarios and participants' motivation level to achieve their goals emerge as key predictors for the overall user experience. Additionally, participants in Exploratory Dialogues and Focused Dialogues perceive the summary as significantly more complete than Structured Choice Dialogues, while participants in Exploratory Dialogues had a significantly lower DEQ compared to those in both the Focused Dialogues and Structured Choice Dialogues conditions.

6.2.1. Interaction between education and strategy on DEQ

Figure 3 depicts the interaction between education levels and dialogue strategy in terms of DEQ. These results suggest that, for the High School education level, being in the Structured Choice Dialogues or Exploratory Dialogues leads to a significant increase in the DEQ mean compared to Focused Dialogues. A plausible interpretation of this observation is grounded in the inherent structure of each strategy. As the education level transits from High School to 'Bachelor and above' education, we see the DEQ mean for Exploratory Dialogues decrease, for Structured Choice Dialogues slightly decrease, for Focused Dialogues increases. This shift suggests that Focused Dialogues might be more suitable or preferred for those with higher education, while Exploratory Dialogues and Structured Choice Dialogues might be more effective for those with a High School education level.

Summary Quantitative Analysis and Real-world Implication

The insights highlight the nuanced interplay between dialogue strategies, self-conscious emotions, and the

richness of user responses. Interestingly, while the study did not reveal significant distinctions in self-conscious emotions such as shame, guilt, and hubristic pride across different strategies, certain trends emerge in relation to authentic pride and user experience.

The Exploratory Dialogues strategy, in particular, registered lower scores for authentic pride and DEQ, while Structured Choice Dialogues showed a noticeable decline in the completeness of the provided information. This suggests a trade-off exists between the level of restraints imposed by the dialogue agent's questions and the likelihood of acquiring necessary information from users. Employing explicit prompts, as seen in the Focused Dialogues or Structured Choice Dialogues approaches, tends to yield more precise information. However, such specificity might constrict the depth or breadth of user elaboration, with the Structured Choice Dialogues method being a prime example. Focused Dialogues, while beneficial for many, can pose interpretative challenges, particularly for individuals with solely a high school education. Contrarily, the unrestricted nature of the Exploratory Dialogues method, though fostering freedom of expression, appears to diminish feelings of pride in respondents.

In essence, each dialogue strategy presents unique advantages and limitations. Real-world implementations must cater to diverse users and contexts. For instance, a hybrid approach could be employed, combining Structured Choice Dialogues as guiding questions followed by open-ended questions to allow users to articulate freely or provide supplementary information. Additionally, Dewdney and Michell (1997) suggests the importance of contextualisation and using neutral phrasing when employing why-questions to mitigate the risk of eliciting unproductive or even hostile responses.

6.3. Text responses comparison between exploratory dialogues and focused dialogues

The deductive analysis aimed to identify and categorise the emergence of novel information related to social contexts, environmental contexts, or personal values within the responses. Utilizing an inductive coding approach, we analyzed responses that introduced emerging topics. This in-depth examination allowed us to understand the breadth of content participants shared during alignment dialogues.

From the results, both the Exploratory Dialogues and Focused Dialogues, across each query round, showed an average of approximately 62% inclination to provide novel information.

It is worth noting that the last 'why' query within the Exploratory Dialogues condition exhibited a mere 27%

introduction of novel information. Such a decline might hint that participants were no longer able to add more reasons by the fourth question or an ambiguous interpretation stemming from the preceding (Q3) query. Another observation of Q3 is its high percentage of novel information elicited. This could be because Q3 is very specific in asking ‘What feelings does it give you.’ Our intention behind leaving Q2 and Q4 (‘Why is it important to you?’) ambiguous was to gauge how participants would respond when presented with very open-ended queries. The significant reduction in novel information suggests there might be a limit to the effectiveness of repeatedly asking open-ended questions. This could be attributed to participants reaching a point where they no longer provide new insights or being fatigued from the continuous questioning.

Within the Exploratory Dialogues condition, the findings showed that on average only 7% of the novel responses touched on social aspects or contextual facets, while a relatively higher 15% on personal values. Such numbers, being on the lower side, suggest the difficulty in sourcing comprehensive data for constructing an user model using very open-ended questioning. In contrast, the Focused Dialogues condition showcased better results: 48% novel responses introduced social nuances, 27% provided context, and 30% on personal values.

An interesting observation from the Exploratory Dialogues data is the higher number of value-laden responses compared to mentions of social or contextual factors. The recurrent theme in Exploratory Dialogues’ questions, centring on ‘why is it important to you,’ might have steered participants to introspect and articulate their intrinsic values more. However, nuanced specifics about the situation, especially those tied to social or environmental contexts, seldom emerged organically from the respondents.

Summary Deductive Analysis and Real-world Implication

Our experiments show that specificity in questions increases the chances of acquiring the targeted information. However, a more restrictive approach, like the multiple-choice format used in the Structured Choice Dialogues, might limit the depth and variety of responses. Real-world implementations need to find a balance between providing structure and allowing flexibility. Combining specific, guiding questions with opportunities for open-ended responses could help gather comprehensive insights while accommodating diverse user needs and contexts.

6.4. New topics for user model

The inductive coding analysis revealed new topics (Section 5.3.2) that can enrich the user model. Below we group similar topics to discuss.

6.4.1. Individual challenges and environmental and societal challenges

A significant finding was the emergence of ‘Individual Challenges’ and ‘Environmental and Societal Challenges.’ These topics underscore the importance of considering various constraints that users may face. These constraints often fall beyond users’ control, such as personal challenges, health conditions, or specific work-related restrictions. Including these in the user model offers a deeper capture of a user’s background and environment. Notably, these findings align with the concept of ‘barriers’ discussed in the behaviour change literature, emphasising factors like family commitments, societal pressures, limited facilities, or perceived time constraints (e.g. Cradock et al. 2021; Skoglund et al. 2022; Sullivan and Lachman 2017).

However, we suggest that such factors should be treated separately from the information elicited in alignment dialogues. Social and contextual aspects are situation-specific, while personal and environmental constraints represent more general background information that could be collected during the user initialisation phase, as discussed in Section 1.

6.4.2. Preferences, norms, and obligations

We also identified a topic revolving around attitudes, preferences, obligations, and the like. This topic delves into users’ preferences for various activities and how they prioritise these activities in their lives, similar to the findings in Kearney’s work (Kearney and McElhone 1999). While some preferences can reflect underlying values, others are simply personal choices devoid of deeper meaning. The related topic, ‘Norms and Obligations’ serves similar functions but instead of personal preference, they are usually driven by societal and personal norms, obligations, and external pressures (Lapinski and Rimal 2005). Whether it is a matter of prioritisation, preference, or adhering to norms, these insights can be integrated as ‘norms’ in the user model, guiding the selection of support actions by the agent.

6.4.3. Tangible outcomes & emotional outcomes

The remaining topics primarily revolved around the outcomes of participants’ actions or behaviours, categorised as desires (typically positive outcomes) and consequences (often negative outcomes). Some of these

outcomes are deeply intertwined with personal values, as depicted by statements like ‘So that I can take care of my beloved ones.’ Conversely, others reflect a more pragmatic stance and do not necessarily embody underlying values, as in the statement ‘I will go to jail for negligence.’ In the Exploratory Dialogues condition, one of the question rounds (Q3) asked ‘What feeling does it give you?’ Such a question yielded valuable insights in terms of emotional consequences. Nearly 80% of responses included novel information, with 2.6% containing topics related to values.

Additionally, subcategories (see Section 5.3.2) – specifically ‘Outcomes about Work and Career/Social Relationships with Others/Appearance’ – are closely intertwined with values like warm relationships with others, security or self-fulfillment, and self-respect. Based on this, we posit that exploring users’ anticipated outcomes (both desires and consequences) could be a strategic method to delve deeper into their core values.

The incorporation of these novel topics into the user model introduces a more holistic and personalised understanding of users, enabling agents to offer personalised support that respects individual constraints, priorities, and motivations. However, it is crucial to strike a balance between the comprehensiveness of the user model and the efficiency of interactions. Careful consideration is needed to avoid overwhelming users or agents with excessive detail.

6.5. Pitfalls of alignment dialogues

Our analysis identified some challenges with alignment dialogues. In Structured Choice Dialogues, we observed instances where participant responses did not logically align with their stated non-adherence scenarios. For example, a participant whose goal is to go to the gym cited ‘contact frequency’ as a contributing factor while being sick—a scenario where social factors seem irrelevant. Such responses suggest a need for deeper user engagement or point to possible misunderstandings of the dialogue prompts.

Furthermore, in Exploratory Dialogues, approximately 15% of the responses were not about why not adhering to goals but about the reasons behind setting these goals in the first place. This underscores the inherent ambiguity in open-ended questioning and the difficulty in asking clear, purpose-driven questions.

Despite these issues, we chose not to exclude these ‘incorrect’ responses from our analysis for two reasons. First, the proportion of these responses was not substantial. Second, all responses, regardless of their direct relevance to non-adherence reasons, contribute valuable

insights into users’ reasons and can enhance our understanding of their behaviour choices.

7. Reflection on alignment dialogue design

In this section, we further discuss the insights from this study and how they may inform the design of alignment dialogues. We begin with an analysis of non-adherence scenarios Section 7.1, as they serve as the starting points for initiating an alignment dialogue. Lastly, we analyze the conversation ‘paths’ by examining the percentages of the sequential occurrence of topics Section 7.2.

7.1. Non-adherence scenarios

We classified the non-adherence scenarios reported by participants during the initial phase of our experiment. Our rationale stems from the understanding that distinct scenarios might require different dialogue strategies to be effective.

Many of the non-adherence scenarios reported by participants align with topics identified in the inductive coding of responses (Section 5.3.2). This overlap was expected, as participants’ explanations expanded upon the initial scenarios, providing additional insights. Nonetheless, we also identified several unique non-adherence scenarios that distinguished themselves from the common categories:

- (1) Lack of Motivation
- (2) Lack of Energy
- (3) Emotional State
- (4) Time constraints

These scenarios primarily highlight participants’ mental or motivational states hindering their commitment to goal behaviours. Participants often provided general reasons without delving into specific circumstances or underlying causes of non-adherence. Vague explanations lack specificity, failing to offer actionable insights into user behaviour. For instance, the statement ‘I don’t have time’ lacks clarity on activities consuming time and potential adjustments. Deeper exploration is needed to uncover underlying factors in such scenarios.

7.2. Dialogue paths for alignment dialogue

Drawing from the Preceding Path Analysis and Following Path Analysis of the duos (detailed in Section 5.4), several patterns emerge that can inform effective dialogue design. Firstly, when participants present vague scenarios discussed in Section 7.1 like ‘I don’t have time’, the agent’s inquiries should aim to extract specific

details. Following discussions on Individual Challenges or Environmental and Societal Challenges, further inquiries often yield Contextual specifics about the situation. This indicates that in real-world implementations, when users discuss Individual Challenges or Environmental and Societal Challenges, it would be effective for the agent to follow up with inquiries about the Contextual specifics of the situation.

Furthermore, discussions about Tangible/Emotional Outcomes, Norms and Obligations, Social aspects, or Contextual aspects often lead to conversations about Values. This pattern indicates that these topics can pave the way for discussions about values.

Overall, the recommended dialogue flow in practical applications might progress from addressing Challenges to exploring Prioritization and Preferences, followed by Tangible Outcomes, Norms and Obligations, Beliefs, Emotional Outcomes, and finally, Values. However, the progression is not linear due to the intertwined nature of these topics, necessitating the agent to adapt its approach based on context and user cues.

8. Limitations and future work

8.1. Emotions in alignment dialogues

This study suggests that self-conscious emotions do not significantly differ across different dialogue strategies. However, more research is needed to understand the precise relationship between dialogue strategies, setup, and the experience of self-conscious emotions.

Furthermore, we did not measure self-conscious emotions before and after the participants engaged in the dialogues. As a result, we could not determine which specific dialogue strategies evoke particular emotional responses. Although it isn't the agent's goal to evoke negative self-conscious emotions, understanding the precise relationship between dialogue strategies and these emotions could further strengthen the design of alignment dialogues. This would ensure that the dialogues are both effective and sensitive to users' emotional states.

Additionally, future studies could benefit from including variations in ethnicity, geographic location, cultural and socioeconomic backgrounds into the analysis. These factors could influence how self-conscious emotions are experienced and expressed during alignment dialogues, providing a more comprehensive understanding of the user experience across different demographics.

Another analysis is to explore the types of health goals participants set and identify any patterns associated with specific goals and the challenges they face.

Understanding these patterns could help tailor dialogue strategies to better support users in achieving their health objectives.

8.2. Alignment dialogue design

While this study addresses the *what* aspect of alignment dialogue (i.e. 'what information should the agent seek in alignment dialogue?'), it provides limited guidance on the *how*. Future work is needed to enrich and refine the dialogue path we suggested. Implementing the dialogue in practice will require a more detailed structure, along with clearly defined relationships between each topic. Our work primarily outlines the topics but does not model how these topics interconnect. Additionally, the relevance of each topic may vary depending on the situation. Identifying which topics are pertinent in specific contexts is a critical area for future exploration.

9. Conclusion

The purpose of the present work was to investigate (1) the effects of different dialogue strategies on users' self-conscious emotions, dialogue experience, perceived accuracy, and perceived completeness (2) which reasons users mention to explain their non-adherence behaviour. The ultimate goal was to better capture *what* constitutes a comprehensive user model and how agents could effectively gather this information without triggering negative emotional responses.

Our findings show that the open-ended general question had significant negative effects on authentic pride, dialogue experience, and the perceived completeness of the explanations collected. For other aspects namely shame, guilt, and perceived accuracy, the dialogue strategies did not show a significant effect.

Furthermore, we used inductive coding to analyze what topics participants mentioned to explain their non-adherence behaviour. This analysis highlighted key topics that should be incorporated into user models. This list of topics is aimed at enabling the agent to form a comprehensive model of users, encompassing their personal constraints, beliefs, values, and preferences. The primary objective of alignment dialogues, therefore, is to gather information within this ontological framework, thus facilitating a deeper and more nuanced understanding of user behaviour in various contexts.

Notes

1. <https://www.prolific.co/>
2. <https://www.qualtrics.com/>
3. <https://landbot.io/>

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work is part of the Hybrid Intelligence Gravitation Programme, with project number 024.004.022, which is financed by the Netherlands Organisation for Scientific Research (NWO).

References

- Abri, Sara, Rayan Abri, and Salih Çetin. 2020. "A Classification on Different Aspects of User Modelling in Alized Web Search." In *Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval*, Toronto, Canada, 194–199.
- Akman, Varol, and Mehmet Surav. 1996. "Steps toward Formalizing Context." *AI Magazine* 17 (3): 55–55.
- Armstrong, Stuart, and Sören Mindermann. 2018. "Occam's Razor is Insufficient to Infer the Preferences of Irrational Agents." In *Advances in Neural Information Processing Systems* 31. Montreal, Canada.
- Ben-Nun, Pazit. 2008. "Respondent Fatigue." In *Encyclopedia of Survey Research Methods*. Vol. 2, 742–743. Sage Publications.
- Berka, Jakub, Jan Balata, Catholijn M. Jonker, Zdenek Mikovec, M. Birna van Riemsdijk, and Myrthe L. Tielman. 2022. "Misalignment in Semantic User Model Elicitation Via Conversational Agents: A Case Study in Navigation Support for Visually Impaired People." *International Journal of Human-Computer Interaction* 38 (18–20): 1909–1925.
- Berkovich, Marina, Jan Marco Leimeister, Axel Hoffmann, and Helmut Krcmar. 2011. "Analysis of Requirements Engineering Techniques for IT-Enabled Product Service Systems." In *2011 Workshop on Requirements Engineering for Systems, Services and Systems-Of-Systems*, 50–58. IEEE.
- Bostyn, Dries H., Sybren Sevenhant, and Arne Roets. 2018. "Of Mice, Men, and Trolleys: Hypothetical Judgment versus Real-Life Behavior in Trolley-Style Moral Dilemmas." *Psychological Science* 29 (7): 1084–1093.
- Brachman, Ronald J. 2004. *Knowledge Representation and Reasoning*. San Francisco: Elsevier.
- Butlin, Patrick. 2021. "AI Alignment and Human Reward." In *Proceedings of the 2021AAAI/ACM Conference on AI, Ethics, and Society. Virtual Conference*, 437–445.
- Carver, Charles S., Sungchoon Sinclair, and Sheri L. Johnson. 2010. "Authentic and Hubristic Pride: Differential Relations to Aspects of Goal Regulation, Affect, and Self-Control." *Journal of Research in Personality* 44 (6): 698–703.
- Castonguay, Andree L., Eva Pila, Carsten Wrosch, and Catherine M. Sabiston. 2015. "Body-Related Self-Conscious Emotions Relate to Physical Activity Motivation and Behavior in Men." *American Journal of Men's Health* 9 (3): 209–221.
- Castonguay, Andrée L., Catherine M. Sabiston, Kent C. Kowalski, and Philip M. Wilson. 2016. "Introducing An Instrument to Measure Body and Fitness-Related Self-Conscious Emotions: The BSE-FIT." *Psychology of Sport and Exercise* 23:1–12. <https://doi.org/10.1016/j.psychsport.2015.10.003>.
- Chen, Yukina. 2017. *The Effects of Question Customization on the Quality of an Open-Ended Question*. Nebraska Department of Education, Data, Research, and Evaluation.
- Chen, Pei-Yu, Myrthe L. Tielman, Dirk K. J. Heylen, Catholijn M. Jonker, and M. Birna Van Riemsdijk. 2023. "Acquiring Semantic Knowledge for User Model Updates via Human-Agent Alignment Dialogues." In *HAI 2023: Augmenting Human Intellect*, 93–107. Munich, Germany: IOS Press.
- Clark, Leigh, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, and Cosmin Munteanu. 2019. "What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. Glasgow, UK.
- Crabtree, B. F., and W. F. Miller. 1992. A Template Approach to Text Analysis: Developing and Using Codebooks. In B. F. Crabtree & W. L. Miller (Eds.), *Doing Qualitative Research* (pp. 93–109). Sage Publications, Inc.
- Craddock, Kevin A., Leo R. Quinlan, Francis M. Finucane, Heather L. Gainforth, Kathleen A. Martin Ginis, Ana Correia de Barros, Elizabeth B. N. Sanders, and Gearóid ÓLaighin. 2021. "Identifying Barriers and Facilitators to Diet and Physical Activity Behaviour Change in Type 2 Diabetes Using a Design Probe Methodology." *Journal of Personalized Medicine* 11 (2): 72. <https://doi.org/10.3390/jpm11020072>.
- Deslippe, Alysha L., Alexandra Soanes, Celeste C. Bouchaud, Hailee Beckenstein, May Slim, Hugues Plourde, and Tamara R. Cohen. 2023. "Barriers and Facilitators to Diet, Physical Activity and Lifestyle Behavior Intervention Adherence: A Qualitative Systematic Review of the Literature." *International Journal of Behavioral Nutrition and Physical Activity* 20 (1): 1–25. <https://doi.org/10.1186/s12966-023-01424-2>.
- Dewdney, Patricia, and Gillian Michell. 1997. "Asking 'why' Questions in the Reference Interview: A Theoretical Justification." *The Library Quarterly* 67 (1): 50–71. <https://doi.org/10.1086/629910>.
- Dignum, Virginia. 2017. "Responsible Autonomy." In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, 4698–4704. Melbourne, Australia. <https://doi.org/10.24963/ijcai.2017/655>.
- Dingler, Tilman, Dominika Kwasnicka, Jing Wei, Enying Gong, and Brian Oldenburg. 2021. "The Use and Promise of Conversational Agents in Digital Health." *Yearbook of Medical Informatics* 30 (01): 191–199. <https://doi.org/10.1055/s-0041-1726510>.
- Divine, Richard L., and Lawrence Lepisto. 2005. "Analysis of the Healthy Lifestyle Consumer." *Journal of Consumer Marketing* 22 (5): 275–283. <https://doi.org/10.1108/07363760510611707>.
- Endsley, Mica R. 1995. "Toward a Theory of Situation Awareness in Dynamic Systems." *Human Factors* 37 (1): 32–64. <https://doi.org/10.1518/001872095779049543>.
- Fereday, Jennifer, and Eimear Muir-Cochrane. 2006. "Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development." *International Journal of*

- Qualitative Methods* 5 (1): 80–92. <https://doi.org/10.1177/160940690600500107>.
- Fishbach, Ayelet, and Ying Zhang. 2008. “Together Or apart: When Goals and Temptations Complement versus Compete.” *Journal of Ality and Social Psychology* 94 (4): 547.
- Friedman, B., P. H. Kahn, A. Borning, A. Hultgren. 2013b. Value Sensitive Design and Information Systems. In N. Doorn, D. Schuurbiers, I. van de Poel, M. Gorman (Eds.), *Early Engagement and New Technologies: Opening Up the Laboratory*. Philosophy of Engineering and Technology, vol 16. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-7844-3_4.
- Friedman, Batya, Peter H. Kahn, Alan Borning, and Alina Hultgren. 2013a. “Value Sensitive Design and Information Systems.” *Early Engagement and New Technologies: Opening Up the Laboratory*, 55–95.
- Griol, David, Zoraida Callejas, Ramón López-Cózar, and Giuseppe Riccardi. 2014. “A Domain-Independent Statistical Methodology for Dialog Management in Spoken Dialog Systems.” *Computer Speech & Language* 28 (3): 743–768. <https://doi.org/10.1016/j.csl.2013.09.002>.
- Hill, Patrick L., and Daniel K. Lapsley. 2009. “Persons and Situations in the Moral Domain.” *Journal of Research in Personality* 43 (2): 245–246. <https://doi.org/10.1016/j.jrp.2008.12.034>.
- Homer, Pamela M., and Lynn R. Kahle. 1988. “A Structural Equation Test of the Value-Attitude-behavior Hierarchy.” *Journal of Personality and Social Psychology* 54 (4): 638–646. <https://doi.org/10.1037/0022-3514.54.4.638>.
- Hurmuz, Marian Z. M., Stephanie M. Jansen-Kosterink, Harm op den Akker, and Hermie J. Hermens. 2020. “User Experience and Potential Health Effects of a Conversational Agent-Based Electronic Health Intervention: Protocol for An Observational Cohort Study.” *JMIR Research Protocols* 9 (4): e16641. <https://doi.org/10.2196/16641>.
- Kahle, Lynn R. 1983. *Social Values and Social Change: Adaptation to Life in America*. New York, NY: Praeger.
- Kaptein, Maurits, Panos Markopoulos, Boris De Ruyter, and Emile Aarts. 2015. “Personalizing Persuasive Technologies: Explicit and Implicit Alization Using Persuasion Profiles.” *International Journal of Human-Computer Studies* 77:38–51. <https://doi.org/10.1016/j.ijhcs.2015.01.004>.
- Kearney, John M., and Sinead McElhone. 1999. “Perceived Barriers in Trying to Eat Healthier—results of a Pan-EU Consumer Attitudinal Survey.” *British Journal of Nutrition* 81 (S1): S133–S137. <https://doi.org/10.1017/S0007114599000987>.
- Klopfenstein, Lorenz Cuno, Saverio Delpriori, Silvia Malatini, and Alessandro Bogliolo. 2017. “The Rise of Bots: A Survey of Conversational Interfaces, Patterns, and Paradigms.” In *Proceedings of the 2017 Conference on Designing Interactive Systems*, 555–565. Edinburgh, UK.
- Kola, Ilir, Catholijn M. Jonker, Myrthe L. Tielman, and M. Birna van Riemsdijk. 2020. “Grouping Situations Based on their Psychological Characteristics Gives Insight into Personal Values.” In *MRC@ECAI*, 17–26. Santiago de Compostela, Spain.
- Kola, Ilir, Catholijn M. Jonker, and M. Birna van Riemsdijk. 2019. “Who’s that?—Social Situation Awareness for Behaviour Support Agents: A Feasibility Study.” In *International Workshop on Engineering Multi-Agent Systems*, 127–151. Springer.
- Kola, Ilir, Pradeep K. Murukannaiah, Catholijn M. Jonker, and M. Birna Van Riemsdijk. 2022. “Toward Social Situation Awareness in Support Agents.” *IEEE Intelligent Systems* 37 (5): 50–58. <https://doi.org/10.1109/MIS.2022.3163625>.
- Krebs, Paul, James O. Prochaska, and Joseph S. Rossi. 2010. “A Meta-Analysis of Computer-Tailored Interventions for Health Behavior Change.” *Preventive Medicine* 51 (3-4): 214–221. <https://doi.org/10.1016/j.ypmed.2010.06.004>.
- Lapinski, Maria Knight, and Rajiv N. Rimal. 2005. “An Explication of Social Norms.” *Communication Theory* 15 (2): 127–147. <https://doi.org/10.1111/comt.2005.15.issue-2>.
- Liscio, Enrico, M. T. van der Meer, L. Cavalcante Siebert, N. Mouter, C. M. Jonker, and Pradeep K. Murukannaiah. 2021. “Axies: Identifying and Evaluating Context-Specific Values.” In *20th International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 799–808. Virtual conference.
- Liscio, Enrico, Michiel van der Meer, Luciano C. Siebert, Catholijn M. Jonker, and Pradeep K. Murukannaiah. 2022. “What Values Should An Agent Align with? An Empirical Comparison of General and Context-Specific Values.” *Autonomous Agents and Multi-Agent Systems* 36 (1): 23. <https://doi.org/10.1007/s10458-022-09550-0>.
- Lustria, Mia Liza A., Seth M. Noar, Juliann Cortese, Stephanie K. Van Stee, Robert L. Glueckauf, and Junga Lee. 2013. “A Meta-Analysis of Web-Delivered Tailored Health Behavior Change Interventions.” *Journal of Health Communication* 18 (9): 1039–1069. <https://doi.org/10.1080/10810730.2013.768727>.
- Martinengo, Laura, Ahmad Ishqi Jabir, Westin Wei Tin Goh, Nicholas Yong Wai Lo, Moon-Ho Ringo Ho, Tobias Kowatsch, Rifat Atun, Susan Michie, and Lorainne Tudor Car. 2022. “Conversational Agents in Health Care: Scoping Review of Their Behavior Change Techniques and Underpinning Theory.” *Journal of Medical Internet Research* 24 (10): e39243. <https://doi.org/10.2196/39243>.
- Miles, Susan, and Gene Rowe. 2008. “The Laddering Technique.” In *Doing Social Psychology Research*, 305–343. <https://doi.org/10.1002/9780470776278.ch13>.
- Milne-Ives, Madison, Caroline de Cock, Ernest Lim, Melissa Harper Shehadeh, Nick de Pennington, Guy Mole, Eduardo Normando, and Edward Meinert. 2020. “The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review.” *Journal of Medical Internet Research* 22 (10): e20346. <https://doi.org/10.2196/20346>.
- op den Akker, Harm, Val Jones, and Hermie Hermens. 2010. “Predicting Feedback Compliance in a Teletreatment Application.” In *2010 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL 2010)*, 1–5. IEEE.
- op den Akker, Harm, Valerie M. Jones, and Hermie J. Hermens. 2014. “Tailoring Real-Time Physical Activity Coaching Systems: A Literature Survey and Model.” *User Modeling and User-Adapted Interaction* 24 (5): 351–392. <https://doi.org/10.1007/s11257-014-9146-y>.
- Orji, Rita, Julita Vassileva, and Regan L. Mandryk. 2014. “Modeling the Efficacy of Persuasive Strategies for Different Gamer Types in Serious Games for Health.” *User Modeling and User-Adapted Interaction* 24 (5): 453–498. <https://doi.org/10.1007/s11257-014-9149-8>.

- Ozkaramanli, Deger, Elif Özcan, and Pieter Desmet. 2017. "Long-Term Goals Or Immediate Desires? Introducing a Toolset for Designing with Self-Control Dilemmas." *The Design Journal* 20 (2): 219–238. <https://doi.org/10.1080/14606925.2017.1272831>.
- Pasotti, Pietro, Catholijn M. Jonker, and M. Birna van Riemsdijk. 2017. "Action Identification Hierarchies for Behaviour Support Agents." In *Third Workshop on Cognitive Knowledge Acquisition and Applications, Cognitum*. International Joint Conference on Artificial Intelligence 2017, Melbourne, Australia.
- Pereira, Juanan, and Óscar Diaz. 2019. "Using Health Chatbots for Behavior Change: A Mapping Study." *Journal of Medical Systems* 43 (5): 1–13. <https://doi.org/10.1007/s10916-019-1237-1>.
- Pommeranz, Alina, Christian Detweiler, Pascal Wiggers, and Catholijn M. Jonker. 2011. "Self-Reflection on AI Values to Support Value-Sensitive Design." In *Proceedings of HCI 2011 The 25th BCS Conference on Human Computer Interaction* 25, 491–496. Newcastle, UK.
- Reja, Urša, Katja Lozar Manfreda, Valentina Hlebec, and Vasja Vehovar. 2003. "Open-Ended Vs. Close-Ended Questions in Web Questionnaires." *Developments in Applied Statistics* 19 (1): 159–177.
- Rietz, Tim, and Alexander Maedche. 2023. "Ladderbot—A Conversational Agent for Human-Like Online Laddering Interviews." *International Journal of Human-Computer Studies* 171:102969. <https://doi.org/10.1016/j.ijhcs.2022.102969>.
- Rugg, Gordon, Malcolm Eva, Atiya Mahmood, Nazia Rehman, Stephanie Andrews, and Sarah Davies. 2002. "Eliciting Information about Organizational Culture Via Laddering." *Information Systems Journal* 12 (3): 215–229. <https://doi.org/10.1046/j.1365-2575.2002.00124.x>.
- Schwartz, Shalom H. 2012. "An Overview of the Schwartz Theory of Basic Values." *Online Readings in Psychology and Culture* 2 (1): 11. <https://doi.org/10.9707/2307-0919.1116>.
- Skoglund, Gyri, Birgitta Blakstad Nilsson, Cecilie Fromholt Olsen, Astrid Bergland, and Gunvor Hilde. 2022. "Facilitators and Barriers for Lifestyle Change in People with Prediabetes: A Meta-Synthesis of Qualitative Studies." *BMC Public Health* 22 (1): 1–27. <https://doi.org/10.1186/s12889-022-12885-8>.
- Sullivan, Alycia N., and Margie E. Lachman. 2017. "Behavior Change with Fitness Technology in Sedentary Adults: A Review of the Evidence for Increasing Physical Activity." *Frontiers in Public Health* 4:289. <https://doi.org/10.3389/fpubh.2016.00289>.
- Tangney, J. P., R. L. Dearing, P. E. Wagner, and R. Gramzow. 2000. *The Test of Self-Conscious Affect (TOSCA-3)*. Fairfax, VA: George Mason University.
- Tangney, June Price, Jeff Stuewig, and Debra J. Mashek. 2007. "Moral Emotions and Moral Behavior." *Annual Review of Psychology* 58 (1): 345–372. <https://doi.org/10.1146/psych.2007.58.issue-1>.
- ter Heijden, Niels, and Willem-Paul Brinkman. 2011. "Design and Evaluation of a Virtual Reality Exposure Therapy System with Automatic Free Speech Interaction." *Journal of CyberTherapy and Rehabilitation* 4 (1): 41–55.
- Thomas, David R. 2003. "A General Inductive Approach for Qualitative Data Analysis."
- Tielman, Myrthe L., Catholijn M. Jonker, and M. Birna van Riemsdijk. 2018. "What Should I Do? Deriving Norms from Actions, Values and Context." In *MRC@IJCAI*, 35–40. Stockholm, Sweden.
- Tracy, Jessica L., and Richard W. Robins. 2004. "Putting the Self into Self-Conscious Emotions: A Theoretical Model." *Psychological Inquiry* 15 (2): 103–125. https://doi.org/10.1207/s15327965pli1502_01.
- Traum, David R., and Elizabeth A. Hinkelman. 1992. "Conversation Acts in Task-Oriented Spoken Dialogue." *Computational Intelligence* 8 (3): 575–599. <https://doi.org/10.1111/coin.1992.8.issue-3>.
- Vanden Abeele, Vero, Erik Hauters, and Bieke Zaman. 2012. "Increasing the Reliability and Validity of Quantitative Laddering Data with LadderUX." In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, 2057–2062: Austin, TX.
- Van den Hoven, Jeroen, Pieter E. Vermaas, and Ibo Van de Poel. 2015. *Handbook of Ethics, Values, and Technological Design*. Springer Dordrecht.
- van de Poel, I. (2013). Translating Values into Design Requirements. In D. Michelfelder, N. McCarthy, D. Goldberg. (Eds.), *Philosophy and Engineering: Reflections on Practice, Principles and Process*. Philosophy of Engineering and Technology, vol 15. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-7762-0_20
- Van Riemsdijk, M. Birna, Catholijn M. Jonker, and Victor Lesser. 2015. "Creating Socially Adaptive Electronic Partners: Interaction, Reasoning and Ethical Challenges." In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, 1201–1206. Citeseer: Istanbul, Turkey.
- Williams, Emily D., Dominique Bird, Andrew W. Forbes, Anthony Russell, Susan Ash, Robert Friedman, Paul A. Scuffham, and Brian Oldenburg. 2012. "Randomised Controlled Trial of An Automated, Interactive Telephone Intervention (TLC Diabetes) to Improve Type 2 Diabetes Management: Baseline Findings and Six-Month Outcomes." *BMC Public Health* 12 (1): 1–11. <https://doi.org/10.1186/1471-2458-12-602>.