Does Tone Matter? Exploring Context-Aware Explanations in Route Recommendations

Marloes Vredenborg
Information and Computing Sciences
Utrecht University
Utrecht, Netherlands
m.t.r.vredenborg@uu.nl

Marit Bentvelzen
Information and Computing Sciences
Utrecht University
Utrecht, Netherlands
M.Bentvelzen@uu.nl

Christine Bauer
Department of Artificial Intelligence and Human Interfaces
University of Salzburg
Salzburg, Austria
christine.bauer@plus.ac.at

Judith Masthoff
Information and Computing Sciences
Utrecht University
Utrecht, Netherlands
j.f.m.masthoff@uu.nl

Abstract

Recommendation algorithms support decision-making, yet their reasoning is often opaque. Explanations can improve user understanding, but the role of tone has received little attention, particularly in relation to situational context. This is especially relevant for route recommendations, where factors such as urgency shape travellers' needs. Using a mixed-method approach (focus groups, $n = 15$; validation study, $n = 32$; online experiment, $n = 150$), we examine how tone is perceived across contextual conditions in public transport. Unlike prior work comparing tone across domains and users, we focus on within-domain, situational context-dependent effects. Results show that situational context affects tone preferences. In urgent situations, humorous and empathetic tones are less appreciated than neutral or authoritative ones, and some tones are more context-sensitive than others. Our findings emphasise considering situational context when designing recommendation explanations, and offer guidance for designers and researchers.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; • **Information systems** → *Presentation of retrieval results.*

Keywords

explanations, tone, experiment, recommender systems

ACM Reference Format:
Marloes Vredenborg, Marit Bentvelzen, Christine Bauer, and Judith Masthoff. 2026. Does Tone Matter? Exploring Context-Aware Explanations in Route Recommendations. In *34th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '26), June 08–11, 2026, Gothenburg, Sweden.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3774935.3806780

1 Introduction

Algorithmic decision-making increasingly relies on complex Artificial Intelligence (AI) models. Often, these systems merely present outcomes, leaving their reasoning opaque. Recommender Systems (RSs) are a prominent example, providing personalised recommendations [1, 46] across domains such as shopping [32], music [8], education [41], and route planning [57]. Despite their widespread use, users may hesitate to follow recommendations, questioning their accuracy, relevance and adequacy of the underlying information, and whether critical information has been overlooked [52].

Prior research shows that integrating explanations into the User Interface (UI) can help users better understand system outputs [16, 22, 37, 52]. Most work has focused on explanation content and presentation, including level of detail [9], explanation styles [16, 29, 52], and their presentation within UIs [19, 22, 29].

Tone is a key aspect of human communication and is adapted to recipients based on factors such as personality or social relationships, even when the underlying content remains unchanged [21, 30]. While research has studied tone in messages [3, 49], voices [43], and dialogues [10, 55], the role of explanation tone has only recently gained attention [38–40]. In recommendation explanations, tone perceptions vary by domain and user: for example, formal tones are preferred for hotel and product recommendations but less so for movies [39], while user traits such as age and personality further shape preferences [39, 40]. Humorous explanations appeal to extroverts but are less favoured by older users, particularly in product recommendations [39].

Beyond domain- and user-dependent effects, we argue that the effectiveness and appropriateness of explanation tones also depend on a user's situational context within a domain, as it shapes needs and expectations. Public transport route recommendation is an especially relevant domain. Although the number of alternatives may be limited compared to domains like music and movies, multimodal journey planning often produces a combinatory range of plausible route permutations. This yields multiple route recommendations, with various attributes, requiring travellers to rapidly weigh trade-offs. Therefore, explanations are important to help users understand why recommended routes match their constraints and preferences. Also, situational factors such as urgency affect travellers' needs [5, 13, 25]. For instance, under time pressure and high cognitive load, clarity and efficiency may matter most, while in low-stakes situations, reassurance or social presence may matter.

To our knowledge, no prior work has examined how within-domain situational factors shape explanation tone perception in RSs while holding the domain fixed. We address this gap by operationalising situational context as situational state variables—urgency and comfort—that are exogenous to tone and studying their effects on tone perception using a mixed-method approach.

Our exploratory research comprises two studies. Study I uses focus groups to identify (1) relevant travel situations related to urgency and comfort (2) user expectations for explanations, and

UMAP '26, Gothenburg, Sweden

ACM ISBN 979-8-4007-2311-7/26/06
https://doi.org/10.1145/3774935.3806780

(3) attitudes and reactions towards explanation tones in route recommendation. Study II builds on these qualitative insights with an experiment examining how neutral, authoritative, humorous, and empathic tones affect user perceptions across situational contexts.

Our main contributions are methodological: We (i) demonstrate that explanations must account for context, tone, and their interplay (ii) provide a generalisable approach for studying and designing explanations that account for context and tone. In addition, we (iii) demonstrate this finding in a specific domain—public transport (iv) provide qualitative insights into users' needs and expectations for explanations and tones in route recommendation, and (v) provide a validated set of explanation tones for route recommendations.

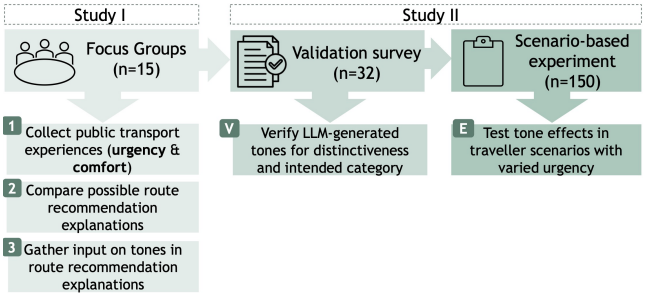


Figure 1: Overview of the two-part study design

2 Related Work

We review related work on explanations in RSs and the role of tone in explanations, with a focus on route recommendation.

Explanations in Recommender Systems. Explanations help users understand, interpret, and justify recommendations [37, 51]. Prior work has identified multiple explanation goals—such as transparency, scrutability, trust, effectiveness, persuasiveness, efficiency, and satisfaction [52]—which often interact or trade off. For example, personalised explanations can increase satisfaction but harm effectiveness [16, 51], transparency can increase satisfaction, and persuasiveness can reduce effectiveness [51].

Approaches to explanations vary in content [22], style [9], and presentation in the UI [19], and can be personalised or adapted to user traits and preferences. While these aspects have been studied across domains such as movies [16, 22], music [29], and products [39], research on explanations' role in route recommendation remains limited [53, 57]. Existing work on explanation in route recommendation often focuses on algorithmic explainability for developers or experts [48], while comparably fewer works propose user-facing or context-aware explanation approaches [26, 45]. Interestingly, specifically in public transport, explanations directed towards travellers [23, 31, 42] rarely account for users' expectations or explanation needs, nor examine how presentation styles shape users' perceptions. As a result, it remains unclear how explanations are beneficial in route recommendation. We address these gaps by investigating what users generally want from route explanations and how different explanation tones are perceived.

Tone in Recommendation Explanations. Tone is a central aspect of human communication, shaping how messages are interpreted and acted upon. People routinely adjust tone to context [36], switching between formal and informal [33], varying emotional support [6], or using humour [34]. In RSs, recent work shows that explanation tone—defined as "*the style or expressiveness of written language, categorising it into styles such as formal or humorous*" [39]—can affect users' perceptions of transparency, trust, effectiveness, and persuasiveness [38]. Notably, this emerging area on tone in RS explanations is driven by a small group of researchers: Okoso et al. [38, 39, 40]. In contrast, the wider field of natural language explanation receives wide attention; see Cambria et al. [7] for an overview and the state of the art in real-time explanation generation.

More concretely, Okoso et al. [38] showed that tone affects explanation goals: attractive, romantic, and formal tones improved perceived transparency, trust, effectiveness, and persuasiveness, while humorous and simple tones showed no consistent benefits. Follow-up work found tone effects to be domain- and user-dependent across movie, hotel, and product RS [39]. Formal tone improved perceptions of hotel and product RSs but not movies, and user traits moderated these effects: extroverted users preferred emotionally rich tones (e.g., attractive, humorous) in movie and hotel domains, while older users reacted negatively to humorous and romantic tones in the product domain. Beyond perception, tone also also affected decision-making depending on the system's perceived role [40]; tones suitable for assistant roles could reduce credibility in expert roles, again moderated by age and extroversion.

Despite this progress, prior work has largely focused on domain- and user-dependent variation, paying little attention to situational context within a single domain. We distinguish domain-dependent effects, arising across application areas (e.g., movies, hotels, products), from context-dependent effects stemming from situational conditions at the moment of interaction. While context may be negligible in some domains, it is critical in public transport, where factors such as time pressure, crowding, and environmental conditions shape travellers' experiences and decision-making [4, 12, 35]. Travellers may shift between urgent commuting, navigating unfamiliar routes, or travelling in low-stress conditions, each potentially requiring different explanation tones. This suggests that explanation tone may need to be adapted not only to users and domains, but also to situational context. We address this gap by systematically investigating how explanation tone interacts with within-domain situational context, holding content constant and varying tone across situational contexts to isolate tone effects in route recommendation. We focus on two contextual dimensions central to public transport experiences: urgency and comfort [5, 13, 14, 18, 25].

3 Overall Study Design

To examine whether explanation tone in route recommendation should adapt to a traveller's situation, we employed a mixed-method research design (see Fig. 1). All studies were allowed to proceed by Utrecht University's Research Institute of Information and Computing Sciences on the basis of an Ethics and Privacy Quick Scan.

In Study I, we explored (1) users' experiences of urgency and comfort in public transport, (2) their expectations for system explanations, and (3) their initial reactions to explanation tones. Beyond providing stand-alone qualitative insights, the findings informed Study II: urgency and comfort experiences were used to create realistic scenarios, and feedback on explanations and tone guided the

creation of explanation styles. Study II employed a scenario-based experiment that held the domain and comfort constant, and manipulated urgency (high vs. low), enabling estimation of the causal effects of situational context on tone perception and the potential of context-aware tone adaptation.

4 Study I: Focus Groups

4.1 Methodology

Participants. We conducted three focus groups with a total of fifteen participants ($n = 15$) recruited through personal networks and snowball sampling. Participants ranged in age from 25 to 42 years ($M = 30.60$, $SD = 4.56$). Eleven identified as female and four as male. All were residents of the European Union. Participant details are available in the Supplementary Material.

Procedure. Each focus group began with a brief introduction and informed consent, followed by three activities (Fig. 1). The full protocol is available in the Supplementary Material.

Activity 1: Urgency and comfort. Participants individually reflected on *urgent* and *non-urgent* public transport experiences, writing descriptions on post-its. They individually mapped these onto a shared low-to-high urgency scale and briefly elaborated on their stories. This was repeated for *comfort*.

Activity 2: Explanations in route recommendations. We introduced route explanations with a travel app example, then showed a regular travel scenario with a route recommendation and five explanation variants differing in level of detail. Participants reviewed them individually, discussed them as a group, and indicated their preferred version. This was repeated for a disrupted scenario.

Activity 3: Tone in route recommendations. Participants individually read the explanation at their group's preferred level of detail, rewritten in several tones, and then discussed their impressions. They indicated which tones they found more or less appropriate, whether any desired tones were missing, and whether preferences differed between regular and disrupted travel.

Materials. We created two fictional public transport scenarios to provide a clear, relatable context. Both described a bus-train journey with specified destinations, timing, and interchanges. One was disruption-free; the other involved a cancelled train. In both cases, the system provided a short route recommendation for continuing the journey. Presenting both regular and disrupted cases helped participants reflect on explanation needs in different situations.

For each route recommendation, we created five explanation levels of increasing detail [9], from minimal to comprehensive, to assess how participants perceived differences in explanation depth. The recommended route was constant; only the explanation text varied. Full explanation texts are in the Supplementary Material.

We also manipulated explanation tone, using categories from Okoso et al. [38, 39, 40] extended with an empathic tone adapted from Saccardi and Masthoff [47]. Empathy is linked to supportive communication [24] and may be relevant during negative experiences like disruptions (see Table 1 for all tones). All tone variants were generated using ChatGPT-4 following Okoso et al.'s procedure [39]; the prompt is provided in the Supplementary Material.

Data Analysis. All focus groups were audio-recorded (5h 35min) and transcribed verbatim. We analysed the data using an iterative, bottom-up coding and synthesis process. For *urgency* and *comfort*, participants' post-it notes were digitized in Microsoft Whiteboard and open-coded to identify relevant experiential elements (e.g., comments about temperature, noise, or stress). Two authors then collaboratively clustered codes using affinity diagramming to derive higher-level categories and themes. Feedback on route recommendation explanations and tones was analysed using the same approach: explanation- and tone-related statements were compiled, iteratively coded, grouped, and synthesised to identify preferences, omissions, and other feedback. Findings across analyses were then integrated to identify overarching themes in participants' perceptions of route recommendation explanations.

Table 1: Overview of tones and characteristics, adapted from Okoso et al. [38, 39] and Saccardi and Masthoff [47].

Tone [A]	Characteristics [B]
Neutral	Objective and factual expression with restrained emotional tone. Simple and direct language
Formal	Polite and sophisticated language. Professional and official tone. Courteous and cautious expression
Authoritative	Confident and assertive language. Persuasive and commanding expression
Casual	Relaxed and informal language. Friendly and conversational tone, often using colloquial expressions
Humorous	Incorporates humour and light-hearted expression. Casual tone with playful language
Romantic	Passionate and emotionally rich expression. Poetic and beautiful language
Attractive	Captivating, engaging, and appealing to the sense or emotions. Vivid adjectives, descriptive language, and storytelling techniques to make the sentence more interesting and engaging
Simple	Simple, clear, and easily understandable by children. Straightforward language, short sentences and avoiding complex vocabulary or concepts. Explaining any unfamiliar terms
Empathic	Warm and supportive language. Acknowledges the user's feeling and offers reassurance or encouragement. Creates a sense of care, connection, and emotional understanding.

4.2 Findings

Experiences of urgency. Participants often associated urgency with fixed commitments, such as flights, international trains, or job interviews, which they saw as difficult to reschedule and therefore created a strong sense of urgency from the start. This sense of urgency sometimes shaped decisions made before travel. One participant, for example, explained that a job interview led them to plan carefully by selecting the mode of transport they perceived as most reliable. However, urgency varied by activity: commuting, classes, or concerts were perceived as either high or low urgency depending on the individual, whereas visits to family or friends were generally seen as low urgency due to their flexibility.

Focus groups further showed that perceived urgency was driven by journey stakes and the consequences of delays or cancellations. Participants often mentioned international flights, trains, job interviews, or important meetings, where missing them meant lost investments or opportunities: *"Missing a flight is expensive"*–FG3-P4.

Others noted practical inconveniences, such as longer journeys, being stranded at night, or waiting in discomfort. When consequences were perceived as minor, urgency was described as much lower, as appointments could be postponed or repeated without major impact. Participants also noted how situational conditions affect feelings of urgency, highlighting the importance of considering

travellers' broader context. For example, a concert felt more urgent not only due to its fixed start time but also because anticipated rush-hour crowds added uncertainty and required planning.

Others linked short transfer times and infrequent buses to heightened urgency, especially during late-night travel when missing the last train was a risk (*"I usually arrive [at the airport] at midnight, and then I have to catch the last train."*–FG2-P6)

Participants noted contextual factors that reduced the feeling of urgency, such as direct trains and high-frequency services which lower transfer risks and provide backup options. Similarly, daytime travel and quiet conditions made the journey more relaxed.

Focus groups showed that emotional responses to situations could amplify or soften the sense of urgency. For example, when people were already in contexts that felt urgent, such as travelling along an unfamiliar route, feelings of uncertainty often intensified that urgency. Similarly, when unexpected changes occurred during a journey, emotions shaped how these were experienced. A lack of clear information during delays often heightened stress and powerlessness, increasing the situation's perceived urgency.

Although many participants experienced stress in uncertain situations, some reframed disruptions as opportunities to relax or make the best of the situation, such as going out for dinner.

Experiences of comfort. Participants noted that both fellow travellers' behaviour and the broader travel context influenced their comfort. Crowding, encroachment on personal space, and disregard for social norms—such as loud conversations, talking in silent compartments, or disruptive behaviour—made journeys uncomfortable or stressful. Safety and a sense of control were major concerns: for instance, being stuck in crowded compartments with aggressive or intoxicated passengers increased discomfort and sometimes fear. As one participant recalled: *"It was a very busy train. [...] people were standing everywhere. And then everyone was drunk, people were smoking inside the train and like, playing with their lighters, which is just very uncomfortable and a bit scary... it felt like it was going to escalate. But it was so busy that there was just nowhere to go."*–FG1-P5

Situations that increased vulnerability, such as travelling late, being alone in unfamiliar stations, or encountering deserted areas, undermined comfort. In contrast, staff presence, such as a conductor passing through the train, provided reassurance.

Beyond social interactions and safety concerns, participants indicated that comfort was shaped by the journey's physical and sensory aspects, such as temperature, space, noise, and bodily sensations. Crowding was a common source of physical and sensory discomfort, including standing for long periods, lack of personal space, unwanted physical contact, and noise.

Temperature further affected comfort. Warm or poorly ventilated vehicles, particularly in summer, caused discomfort, often worsened by sweaty or smelly passengers during crowding. As one participant noted: *"During peak hours, when everyone is returning or going to work. The weather was super warm [...] there was no air conditioning, no ventilation. Everyone was sweaty, smelling, and standing because it was like overcrowded... this is like the worst thing for me."*–FG2-P3. Conversely, well-regulated temperature, new vehicles, and a clean, fresh environment were noted to enhance comfort.

While social interactions, safety, and the physical environment strongly influenced comfort, practical aspects also mattered. Amenities like charging points, clear signage, and timely announcements improved comfort, while missing or inconsistent information, poorly maintained vehicles, or inadequate services reduced it.

Overall, participants' comfort was closely tied to their expectations: meeting or exceeding them increased comfort, while falling short led to discomfort. Expectations varied and were influenced by prior experiences with other transport systems or contexts.

Attitudes towards explanations. Participants did not always see the immediate value of explanations for route recommendations, assuming the suggested route was simply the fastest, so the "why" seemed obvious and did not require clarification. As a result, explanations seemed unnecessary in regular situations or even distracting. Instead, participants wanted explanations when journeys did not go as planned. In those situations, they sought reassurance that the system was aware of the problem and was using up-to-date information: *"If it's normal, I don't need it. But if something goes wrong, then I do want to know why you give me this route."*–FG3-P2.

Participants emphasised clarity and conciseness, preferring short, practical explanations that foreground key information. Long or repetitive text was dismissed as *"a lot of words, saying little"* (FG3-P4), while overly sparse explanations were also unhelpful. Explanations should avoid repeating information already provided in the route recommendation, balancing usefulness and brevity. Bullet points were suggested as an effective, scannable format.

Explanations became more meaningful when highlighting comparisons and trade-offs, helping participants understand not only why a route was chosen, but also how it differed from alternatives. For example, some would choose a longer or less direct route to reduce transfers, increase comfort, or avoid crowds: *"I would read the details if there were two options side by side—then you can decide, faster with more walking or slower but shorter"*–FG1-P1.

Finally, explanations were considered most useful when *contextualised*, explicitly referring to the current situation, which signalled awareness of real-time conditions: *"It's nice when it starts with: the train to [city] is cancelled, and then you know the new situation is based on that"*–FG3-P2. Some participants wanted explanations for disruptions (e.g., why a train is cancelled or the likelihood that the next one runs as planned). However, not all context was considered relevant, e.g., external factors like weather were only considered helpful if directly affecting the travel experience. Overall, participants wanted explanations that use context to make reasoning transparent, as it helps judge whether the recommendation makes sense and clarifies why certain factors influenced the decision.

Attitudes towards different tones. Participants had mixed reactions to explanation tones in route recommendations. Many appreciated an authoritative tone for its directness and familiarity, resembling official travel communication and conveying certainty and structure: *"I like that it starts with: this is the optimal route. Let's go."*–FG1-P3. However, others found it overly prescriptive, limiting their sense of agency: *"It's like 'making it a clear choice'. [..] well, no, I will decide what the clear choice [is]."*–FG2-P1. In contrast, softer or hedging language (e.g.,"we think") undermined trust. Some participants perceived neutral, formal, and authoritative tones as similar and generally easiest to follow. Brevity was valued: *"It is short and concise. [..] I just want quick information."*–FG2-P6. Yet, others felt this simplicity was 'too straightforward'.

Participants across all focus groups rejected empathic or emotional tones. They argued that a system cannot know how they feel and should not act as if it does: *"It is not human. This one acts as if the system is a person. I find that really annoying, because it isn't—it's*

just my phone."–FG3-P1, and *"I feel like if you actually see the person telling you this, then it's a different thing rather than reading it off a screen. So if a person tells you [empathic text] with a smile, you're like, okay, that's fine. But if it's written there, you're like no"*–FG1-P1. Nevertheless, some mentioned that an empathic closing could be comforting in situations where they had no control. Some assumed that older users might appreciate a warmer tone more.

Casual tones also drew mixed reactions: many found them inappropriate in stressful situations, while others enjoyed a less formal style once essential information was clear. The attractive tone was described as "sales language", which felt out of place. Yet, one participant emphasised that tone perception may differ if the message were spoken rather than written: *I think it would really matter if it's a text recommendation or like a spoken one [..]. It might matter what kind of actual voice is saying this.*–FG2-P1.

Humorous tones were similarly divisive: some found them inappropriate when rushed or worried, while others felt they eased stress. However, several participants noted that humour worked better when coming from people rather than systems. For both casual and humorous tones, some noted that they may suit leisure travel better, when people are more open to playful messaging.

The romantic tone was least appreciated, described as overly dramatic or "like a computer game", and not fitting the context of everyday travel. However, one participant noted this may reflect unfamiliarity with such tones: *"because we're not used to that kind of information in that way. That maybe it feels a bit weird."*–FG1-P2. This suggests that more formal tones may have felt less surprising because they matched expectations of travel communication.

Overall, participants reacted very differently to tone. Authoritative and simple styles were most often trusted, provided they avoided hedging or exaggerated slogans. Empathic, casual, humorous, and attractive tones were more divisive, with some finding them playful or comforting, while others considered them irritating or misplaced. Romantic language was consistently rejected. These results suggest that tone effectiveness depends strongly on context and personal preference, rather than one style working universally.

5 Study II: Experiment

Study II examines how explanation tone affects user perceptions across contexts, focusing on Urgency in route recommendation.

5.1 Experiment Design

To examine tone effects, we conducted an online experiment informed by Study I, using a mixed design with Urgency (*high* vs. *low*) as within-subjects factor and Tone as between-subjects factor. Holding the domain fixed (public transport route recommendation), we manipulated situational urgency via two travel scenarios informed by Study I: a high-urgency scenario with time pressure and potential negative consequences, and a low-urgency scenario with minimal consequences and no time constraints. To keep the design manageable and avoid confounding effects, Comfort was held constant. In both scenarios, a disruption triggered a route recommendation with an explanatory text varying by tone.

Tone had four between-subjects levels (*neutral*, *authoritative*, *humorous*, *empathic*). The tone selection was informed by Study I. Participants were randomly assigned to one tone and viewed only explanations written in that tone. The explanation's informational content was identical across tone conditions.

Before the main experiment, we conducted a tone validation to ensure explanations accurately conveyed the intended tones and that any effects observed in the main study could be attributed to differences in tone, rather than ambiguities in interpretation. We first describe this validation (Section 5.2), followed by participants (Section 5.3), procedure (Section 5.4), and results (Section 5.5).

5.2 Tone Validation

We recruited $n = 32$ participants using Prolific[1] and reimbursed them with £1.20[2]. Participants were required to be fluent in English and pass two attention checks. Ages ranged from 20 to 65 years ($M = 38.90$, $SD = 12.38$), with equal gender representation (16 male, 16 female). The sample was internationally diverse: UK (55%), Canada (16%), South Africa (13%), and Italy (7%), with one participant each from India, the United States, and Algeria (see Supplementary Material for details).

We conducted the validation using Qualtrics[3]. After providing consent, participants read a hypothetical scenario, in which *Sam*, travelling to visit a friend, faced a train cancellation and received an alternative route recommendation from a route recommendation. The recommendation reflected focus group-identified preferences for fastest arrival, minimal interchanges, and uncrowded trains. Participants then read an explanation of the recommended route, designed based on focus group insights favouring concise, structured comparisons of options. This informed the creation of a *baseline explanation* (available in the Supplementary Material.

Participants evaluated a series of short explanation texts describing why a particular public transport route was recommended. We created tone variants of this explanation using GPT-5 (ChatGPT). We prompted the model with the baseline text and instructed it to rewrite the explanation according to a set of predefined tones while preserving the factual content. The specific prompt used to generate the tone variants is available in the Supplementary Material.

We generated five explanation versions for each tone using this prompt. Participants viewed all 20 explanation texts in randomised order. For each text, they selected the tone they believed it represented from a predefined list of tone categories. An *Other* option was available if none of the listed categories seemed appropriate.

To assess whether the generated statements reflected the intended tone and identify which variant best represented the tone, we measured participant agreement using Free-Marginal Kappa [44], where 0.41–0.60 = moderate, 0.61–0.80 = good, and >= 0.81 = very good agreement. All tone variants reached at least moderate agreement, with the highest-performing variants, Neutral_v4 ($\kappa = 0.57$), Authoritative_v5 ($\kappa = 0.63$), Humorous_v5 ($\kappa = 0.85$), and Empathic_v4 ($\kappa = 0.75$), showing moderate to very good agreement and were selected for the experiment. The texts and free-marginal kappa's for all tones are available in the Supplementary Material.

Fig. 2 shows the texts of the final tones that performed best and were used in the experiment.

[1] https://www.prolific.com/
[2] Equivalent to an hourly rate of £9
[3] https://www.qualtrics.com/

Neutral

Your planned connecting train is cancelled due to a problem on the line. Normal services will resume in about 1 hour. This route is recommended as the most effective choice.

Compared with the other travel options, it:
+ Gets you there the fastest
+ Involves the fewest changes
- The train may be more full, but this is the trade-off for fewer changes and quicker arrival

Authoritative

Your connecting train is cancelled due to a fault. Full service will take 1 hour to return. Use this route now. It is the right choice.

This route:
+ Brings you to your destination first
+ Needs the fewest interchanges
- The train will be busy, but this is the fair trade for quicker travel with fewer changes.

Humorous

Plot twist: your connecting train is cancelled, all thanks to tech gremlins. Normal trains return in an hour. This route is the champ for now.

It stands out because:
+ Quickest arrival
+ Fewer changes
- Sure, it's busy, but think of it as the VIP express for people who dislike waiting.

Empathic

Sadly, your connecting train has been cancelled due to a technical issue. We understand this may cause worry, and we want to help you find the best way forward. Services should recover in about 1 hour. For now, please take this route.

Compared to other options, this route:
+ Gets you there the earliest
+ Has the fewest interchanges
- It may feel a little busy, but it is the smoother and quicker option.

Figure 2: Texts for the experiment's explanation tones. Content (facts, comparisons, trade-off) is constant, only tone varies.

5.3 Participants

We recruited $n = 150$ participants using Prolific[3] and reimbursed them with £1.20[2]. Participants had to speak English fluently and not have participated in the validation study. Ages ranged 20–78 ($M = 40.47$, $SD = 12.65$). 75 participants identified as male and 75 as female. They came from diverse continents: 64 Europeans, 49 North Americans, 31 Africans, 5 Asians, and 1 South American.

5.4 Procedure

We conducted the experiment using Qualtrics[3]. After a brief introduction and consent, participants completed a short personality questionnaire [17]. They were introduced to Sam, *a public transport user who values efficiency and comfort while travelling. They prefer to arrive quickly, avoid interchanges, and travel in vehicles that are not crowded.* Participants were randomly assigned to a low- or high URGENCY scenario. After reading the scenario, they answered two comprehension questions. Next, they were randomly assigned to one of four tone conditions, imagined themselves in Sam's situation, read an explanation of why a route was recommended, and rated the perceived quality of the explanation and their perception of the travel disruption. Explanation quality was measured using six Likert-type items (1 = strongly disagree, 7 = strongly agree), each corresponding to one of six explanation goals [52]: transparency, trust, efficiency, effectiveness, persuasiveness, and satisfaction. Finally, participants completed the same task for the opposite URGENCY scenario with the same tone. The full survey, including the scenarios, is available in the Supplementary Material.

5.5 Results

We first examine the main effect of URGENCY, followed by the effects of TONE, and finally the interaction between URGENCY and TONE.

We conducted a two-way repeated measures MANOVA to investigate the effects of urgency and tone on explanation qualities. Before analysis, we examined all assumptions underlying the repeated measures MANOVA. Although multivariate normality was violated, MANOVA is robust to such violations when group sizes are fairly equal and exceed 30 participants [2, 50]. Additionally, Pillai's Trace was used for its robustness. For the table with full results, see Supplementary Material.

Main effect urgency.

The two-way repeated measures MANOVA showed a significant effect of URGENCY on the explanation metrics ($F(6, 141) = 3.91$, $p = .001$, Pillai's Trace = .14, $\eta_p^2 = .14$), indicating a large effect of URGENCY on participants' perceptions.

Follow-up univariate tests showed a significant main effect of URGENCY on all explanation metrics: Transparency ($F(1, 146) = 7.44$, $p = .007$, $\eta^2 = .05$), Trustworthiness ($F = 8.45$, $p = .004$, $\eta_p^2 = .06$), Effectiveness ($F = 10.59$, $p = .001$, $\eta_p^2 = .07$), Persuasiveness ($F = 4.42$, $p = .037$, $\eta_p^2 = .03$), Efficiency ($F = 12.65$, $p < .001$, $\eta_p^2 = .08$), and Satisfaction ($F = 10.52$, $p = .001$, $\eta_p^2 = .07$).

Post-hoc Bonferroni-adjusted comparisons[4] showed that HIGH-URGENCY led to significantly lower ratings than LOW-URGENCY across all measures (Fig. 3): Transparency ($M = 5.58$ vs. 5.75, $p = .007$), Trust (4.92 vs. 5.17, $p = .004$), Effectiveness (5.42 vs. 5.71, $p = .001$), Persuasiveness (5.16 vs. 5.43, $p = .037$), Efficiency (5.37 vs. 5.72, $p < .001$), and Satisfaction (4.86 vs. 5.19, p=.001).

Main effect tone. A two-way repeated-measures MANOVA showed a significant effect of TONE on the explanation qualities ($F(18, 429) = 1.82$, $p = .021$, Pillai's Trace = .21, $\eta_p^2 = .07$), indicating a medium effect of TONE on participants' perceptions.

Univariate tests showed significant TONE effects on Trustworthiness ($F(3, 146) = 4.15$, $p = .007$, $\eta_p^2 = .08$), Persuasiveness ($F(3, 146) = 3.31$, $p = 0.022$, $\eta_p^2 = .06$), and Satisfaction ($F(3, 146) = 6.04$, $p < .001$, $\eta_p^2 = .11$). No significant effects were found for Transparency, Effectiveness, or Efficiency (all $p > .05$).

Bonferroni-adjusted post-hoc comparisons showed that AUTHORITATIVE was rated significantly higher than HUMOROUS for Trustworthiness ($M = 5.42$ vs. 4.54, $p = .005$) and Persuasiveness (5.69 vs. 4.88, $p = .013$). For Satisfaction, NEUTRAL ($M = 5.26$), AUTHORITATIVE (5.28), and EMPATHIC (5.31) were all rated higher than HUMOROUS (4.24; all $p \leq .005$). Fig. 4 shows that the HUMOROUS TONE consistently received the lowest ratings across all explanation qualities. It also shows that NEUTRAL and AUTHORITATIVE were generally rated slightly higher than EMPATHIC across the explanation qualities, though differences were not statistically significant.

As the informational content was constant across tone, these differences can be attributed to tone rather than content.

Interaction effects. The two-way repeated-measures MANOVA revealed no statistically significant interaction effect between URGENCY and TONE on the explanation qualities ($F(18, 429) = 0.80$,

[4]When Bonferroni correction is used, the adjusted p-values are reported.

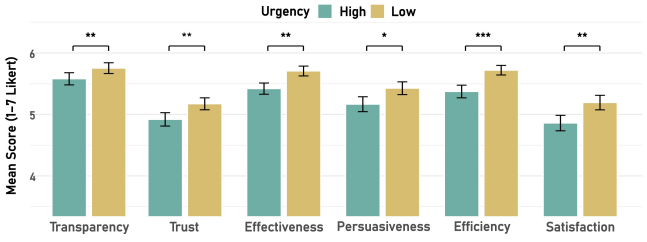


Figure 3: Perceived qualities under high vs. low urgency across dimensions (7-point Likert). $^*p$ <.05, **<.01, ***<.001

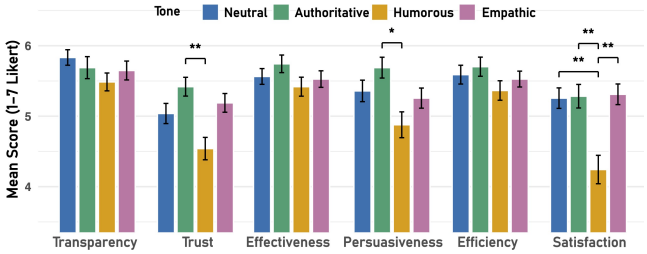


Figure 4: Perceived explanation qualities across four tones and six dimensions (7-point Likert). * $p < .05$, ** $p < .01$

$p = .696$, Pillai's Trace = .10, $\eta_p^2 = .03$). This suggests that the combined influence of urgency and tone did not significantly affect participants' perceptions of the explanation metrics.

Although the overall interaction was not statistically significant, we conducted follow-up analyses. Focus group insights suggested that urgency and tone may not interact uniformly, with specific tone–urgency combinations perceived differently by context. Moreover, given the multidimensional nature of the dependent variables, subtle interaction effects on individual measures may have been obscured in the multivariate test. Therefore, we performed exploratory Bonferroni-corrected pairwise comparisons to identify tone- and urgency-specific patterns across explanation metrics.

To compare tones within each urgency condition, pairwise comparisons were conducted separately for LOW- and HIGH-URGENCY. For Trustworthiness, AUTHORITATIVE was rated significantly higher than HUMOROUS under both HIGH URGENCY ($M = 5.32$ vs. 4.32, $p = .007$), and LOW URGENCY ($p = .037$). For Persuasiveness, under HIGH URGENCY, AUTHORITATIVE was rated significantly higher than HUMOROUS ($M = 5.60$ vs. 4.65, $p = .035$). For Satisfaction, under HIGH URGENCY, HUMOROUS was rated significantly lower than NEUTRAL, AUTHORITATIVE, and EMPATHIC ($M = 4.00$ vs. 5.10, $p = .008$; vs. 5.27, $p = .002$; vs. 5.05, $p = .015$). Under LOW URGENCY, HUMOROUS was rated significantly lower than NEUTRAL and EMPATHIC ($M = 4.49$ vs. 5.41, $p = .028$; vs. 5.57, $p = .007$). No other pairwise differences were significant.

To assess how URGENCY influenced ratings for each TONE, post-hoc comparisons were conducted between HIGH-URGENCY and LOW-URGENCY within each tone. These showed that LOW URGENCY generally resulted in higher ratings than HIGH URGENCY across several measures. For the HUMOROUS TONE, LOW URGENCY resulted in significantly higher Transparency ($M_{low} = 5.62$ vs. $M_{high} = 5.35$, $p = .037$), Trustworthiness (4.76 vs. 4.32, $p = .016$), and Satisfaction (4.49 vs. 4.00, $p = .020$). For the EMPATHIC TONE, LOW URGENCY yielded significantly higher ratings for Effectiveness (5.78 vs. 5.27, $p = .005$), Persuasiveness (5.54 vs. 4.97, $p = .027$), Efficiency (5.73 vs. 5.32, $p = .041$), and Satisfaction (5.57 vs. 5.05, $p = .014$).

Overall, the interaction between tone and urgency was most evident for Humorous and Empathic tones, which differed more between low- and high-urgency. In contrast, Neutral and Authoritative tones showed similar ratings across urgency levels, suggesting little impact of urgency on tone perception. The small yet consistent variations across tones suggest that some tones are more sensitive to context than others. This uneven pattern across tones may help explain the lack of a statistically significant overall interaction effect in the multivariate analysis. As shown in Fig. 5, these tone-specific differences appear as slightly larger gaps between the low- and high-urgency profiles for the Humorous and Empathic tones and near overlap for Neutral and Authoritative tones, highlighting the need for further study of how specific tones are affected by context.

6 Discussion

This work explores how situational context affects users' perceptions of explanation tones in route recommendation. Using a mixed-methods approach, we combined qualitative insights into users' expectations with a quantitative analysis of tone perceptions across contextual scenarios. We discuss our findings below.

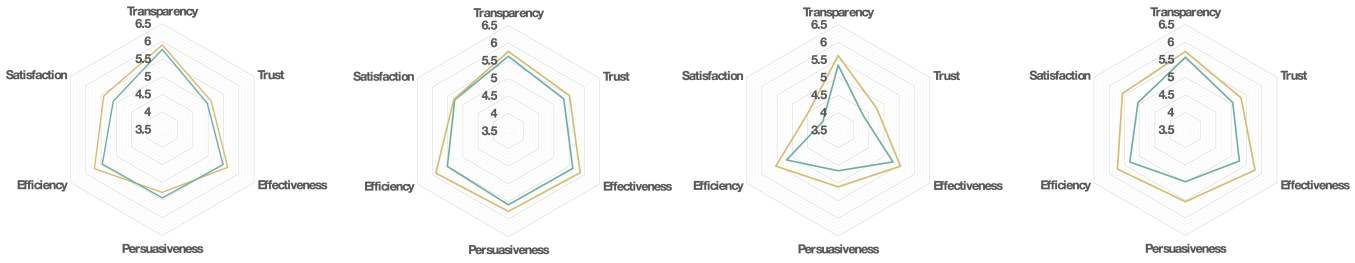
Context and tone affect perceptions of explanations. Our results show that situational context influences how users perceive explanations, even when tone and content remain constant. Explanations were rated more positively in low-urgency situations, whereas in high-urgency scenarios they were consistently rated lower across all explanation qualities. This suggests that perceived explanation quality depends on the context in which it is presented rather than being an inherent property of the explanation.

By varying tone while keeping content constant, our study further shows that tone independently affects perceived explanation quality. A humorous tone consistently received the lowest ratings—particularly for trust, persuasiveness, and satisfaction—indicating it is generally ill-suited for route recommendation. In contrast, an authoritative tone scored highest for trust, persuasiveness, and satisfaction, suggesting that a confident and assertive style enhances perceptions. Focus group findings support that clear, concise explanations make situations transparent and outline actionable steps.

At the same time, qualitative findings revealed individual differences in tone preferences. Some participants perceived the authoritative tone as overly prescriptive, diminishing their sense of agency; others noted that humour might be appreciated in difficult situations. These findings align with prior work [39, 40], indicating that the effects of tone may depend on personal characteristics. Future work should therefore investigate how personal factors interact with situational context to shape tone preferences.

While it may not surprise that a humorous tone is generally less appreciated in route recommendation, our findings reveal a more nuanced picture: tone effects vary systematically with situational context. Some tones appeared more context-sensitive than others. While neutral and authoritative tones were perceived fairly consistently across urgency levels, humorous and empathic tones varied more with context. In low-urgency situations, the humorous tone—and to a lesser extent the empathic tone—were evaluated less negatively than in high-urgency contexts, even though empathy is often associated with supportive communication [24]. Focus group participants suggested that empathetic explanations could feel misplaced or patronising. One possible explanation is that humorous and empathic tones were perceived as less direct, conflicting with preferences for concise communication in high-urgency situations.

Together, these findings show that effective explanation design must consider situational context alongside personal characteristics and domain. Although some tone effects may seem intuitive, this area remains underresearched; documenting and validating tone–context effects is therefore valuable for cross-domain comparisons.



(a) Tone: Neutral (b) Tone: Authoritative (c) Tone: Humorous (d) Tone: Empathic

Figure 5: Mean ratings of explanation quality across four tones, comparing low-urgency (—) and high-urgency (—) conditions.

More broadly, we presented a methodological approach to studying context and tone effects in explanations, with potential relevance to domains such as news recommendation. Future work should extend this approach across domains and further examine how personal characteristics interact with context in shaping tone preferences.

Limits of Humanising RSs. Our findings suggest that human-like tones may evoke unwelcome anthropomorphic perceptions. Several participants felt uncomfortable with humorous, romantic, attractive, and empathetic tones, noting that route systems "cannot feel" and should not mimic human emotions or jokes. While such tones may be acceptable or appreciated in human interaction, many found them inappropriate for route recommendation. This echoes prior work showing that anthropomorphic cues in human–AI interaction can increase trust [56] and warmth [27], but decrease liking [28]. These results raise broader questions about the suitability of human-like qualities in RSs and their explanations. Prior research finds that perceived anthropomorphism can increase trust in RSs but may also reduce willingness to share personal data [54]. Our findings extend this by suggesting that the appropriateness of anthropomorphism—specifically explanation tone—depends on domain and context. Future work should explore the appropriate degree and limits of anthropomorphism in RSs explanations.

Suggestions. Our insights yield five suggestions for RSs designers on using context and tone. First, **adapt explanation tone to situational context rather than treating it as a fixed system characteristic (S1)**. We illustrated this with urgency as the situational context. In other domains, relevant context may differ, so designers should identify and adjust tone to the contextual factors most relevant to their systems. Second, **ensure tone does not compromise clarity, conciseness, or actionability (S2).** Across contexts, users prioritised explanations that clearly communicated trade-offs and reasoning; so, tones should remain secondary to informational clarity. Third, **balance authority with user agency (S3)**. Authoritative explanations can increase trust and persuasiveness, but overly prescriptive phrasing can undermine users' sense of autonomy, especially in non-urgent situations. Fourth, **carefully integrate anthropomorphic characteristics (S4)**, as human-like tones were often seen as inappropriate; users expect route recommendations to be primarily functional. Lastly, **avoid empathic or humorous tones in urgent situations (S5)**. In high-urgency contexts, these tones were rated lower across explanation qualities and often perceived as distracting or inappropriate compared to neutral or authoritative ones.

7 Conclusion

This work offers a first exploration of how context shapes perceptions of explanation tones in route recommendations. Using a mixed-methods approach, we identified explanation needs and assessed responses to tones across contextual scenarios. Results show that tone clearly affects users' perceptions. Beyond prior domain-dependent work, we show that tone effects also vary within a single domain depending on context. In high-urgency situations, users judged tones more critically: humorous and empathic tones were rated lower, while authoritative tones increased trust and satisfaction. Pairwise comparisons further confirmed that tone effects differ by context. Methodologically, we offer a generalisable approach for studying explanations that account for both context and tone, along with a validated set of tones for route recommendations. Overall, our findings highlight that tone does not operate independently from context, and we hope this work encourages further research and practical operationalisation.

We acknowledge several limitations. First, scenario-based stimuli may not fully capture real-world reactions [11, 15], though prior work suggests they can approximate real-world behaviour [20]. To strengthen validity, we used standardised, focus group–informed scenarios and quantitative–qualitative triangulation. Still, scenario presentation may affect tone perception; future work should explore alternative modalities, such as spoken explanations. Second, while our design isolated urgency effects by holding comfort constant, comfort may also influence tone perception. Future research should study interactions between comfort, urgency, and other contextual factors. Third, the non-significant tone–urgency interaction may reflect limited sensitivity to small effects rather than the absence of an interaction. A sensitivity analysis showed that, with $\alpha = .05$ and power = .80, the repeated-measures MANOVA was sensitive to effects of approximately $f = 0.27$ ($\eta_p^2 \approx 0.07$). The study was therefore adequately powered to detect medium-sized interaction effects, but smaller effects may have gone undetected. Future studies should examine these interactions with larger samples. Finally, our samples were geographically diverse: although the focus groups included only European residents, the validation and experiment included participants from multiple continents. This supports broader applicability but adds variability in prior experiences and limits fine-grained cultural comparisons. We therefore encourage further research to validate results in specific regional contexts.

Acknowledgments

We used ChatGPT (v4, v5) to adapt explanation texts to different tones and reviewed all outputs for accuracy and appropriateness. This publication was partially funded by the Dutch Railways. This publication was also supported by the Excellence in Digital Sciences and Interdisciplinary Technologies (EXDIGIT) project, funded by Land Salzburg under grant number 20204-WISS/263/6-6022.

References

[1] Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering* 17, 6 (2005), 734–749. doi:10.1109/TKDE.2005.99

[2] Peter Allen, Kellie Bennett, and Brody Heritage. 2014. *SPSS Statistics Version 22: A Practical Guide.* Cengage Learning, Victoria, Australia.

[3] Sima Ipakchian Askari, Antal Haans, Pieter Bos, Maureen Eggink, Emily Mengfei Lu, Fenella Kwong, and Wijnand IJsselsteijn. 2020. Context Matters: The Effect of Textual Tone on the Evaluation of Mediated Social Touch. In *Haptics: Science, Technology, Applications*, Ilana Nisky, Jess Hartcher-O'Brien, Micha"el Wiertlewski, and Jeroen Smeets (Eds.). Springer International Publishing, Cham, Switzerland, 131–139. doi:10.1007/978-3-030-58147-3_15

[4] Gabriela Beirão and JA Sarsfield Cabral. 2007. Understanding Attitudes Towards Public Transport and Private Car: A Qualitative Study. *Transport Policy* 14, 6 (2007), 478–489. doi:10.1016/j.tranpol.2007.04.009

[5] Candace Brakewood, Sean Barbeau, and Kari Watkins. 2014. An Experiment Evaluating the Impacts of Real-Time Transit Information on Bus Riders inTampa, Florida. *Transportation Research Part A: Policy and Practice* 69 (2014), 409–422. doi:10.1016/j.tra.2014.09.003

[6] Brant R Burleson. 2003. Emotional Support Skills. In *Handbook of Communication and Social Interaction Skills*. Routledge, New York, NY, USA, Chapter 14, 551–594.

[7] Erik Cambria, Lorenzo Malandri, Fabio Mercorio, Mario Mezzanzanica, and Navid Nobani. 2023. A Survey on XAI and Natural Language Explanations. *Information Processing & Management* 60, 1, Article 103111 (Jan. 2023), 16 pages. doi:10.1016/j.ipm.2022.103111

[8] Oscar Celma. 2010. Music recommendation. In *Music Recommendation and Discovery: The Long Tail, Long Fail, and Long Play in the Digital Music Space*. Springer, Berlin, Heidelberg, Germany, 43–85. doi:10.1007/978-3-642-13287-2_3

[9] Mohamed Amine Chatti, Mouadh Guesmi, Laura Vorgerd, Thao Ngo, Shoeb Joarder, Qurat Ul Ain, and Arham Muslim. 2022. Is More Always Better? The Effects of Personal Characteristics and Level of Detail on the Perception of Explanations in a Recommender System. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (Barcelona, Spain) *(UMAP '22)*. Association for Computing Machinery, New York, NY, USA, 254–264. doi:10.1145/3503252.3531304

[10] Ana Paula Chaves, Eck Doerry, Jesse Egbert, and Marco Gerosa. 2019. It's How You Say It: Identifying Appropriate Register for Chatbot Language Design. In *Proceedings of the 7th International Conference on Human-Agent Interaction* (Kyoto, Japan) *(HAI '19)*. Association for Computing Machinery, New York, NY, USA, 102–109. doi:10.1145/3349537.3351901

[11] Jessica L Collett and Ellen Childs. 2011. Minding the Gap: Meaning, Affect, and the Potential Shortcomings of Vignettes. *Social Science Research* 40, 2 (2011), 513–522. doi:10.1016/j.ssresearch.2010.08.008

[12] Luigi Dell'Olio, Angel Ibeas, and Patricia Cecin. 2011. The Quality of Service Desired by Public Transport Users. *Transport Policy* 18, 1 (2011), 217–227. doi:10.1016/j.tranpol.2010.08.005

[13] Arkadiusz Drabicki, Oded Cats, Rafał Kucharski, Achille Fonzone, and Andrzej Szarata. 2023. Should I Stay or Should I Board? Willingness to Wait With Real-Time Crowding Information in Urban Public Transport. *Research in Transportation Business & Management* 47, Article 100963 (2023), 13 pages. doi:10.1016/j.rtbm.2023.100963

[14] Kaja Solland Egset and Trond Nordfjærn. 2019. The Role of Transport Priorities, Transport Attitudes and Situational Factors for Sustainable Transport Mode Use in Wintertime. *Transportation Research Part F: Traffic Psychology and Behaviour* 62 (2019), 473–482. doi:10.1016/j.trf.2019.02.003

[15] Jacquelyn H Flaskerud. 1979. Use of Vignettes to Elicit Responses Toward Broad Concepts. *Nursing Research* 28, 4 (1979), 210–211.

[16] Fatih Gedikli, Dietmar Jannach, and Mouzhi Ge. 2014. How Should I Explain? A Comparison Of Different Explanation Types for Recommender Systems. *International Journal of Human-Computer Studies* 72, 4 (2014), 367–382. doi:10.1016/j.ijhcs.2013.12.007

[17] Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. 2003. A Very Brief Measure of the Big-Five Personality Domains. *Journal of Research in Personality* 37, 6 (2003), 504–528. doi:10.1016/S0092-6566(03)00046-1

[18] Elise Grison, Jean-Marie Burkhardt, and Valérie Gyselinck. 2017. How Do Users Choose Their Routes in Public Transport? The Effect of Individual Profile and Contextual Factors. *Transportation Research Part F: Traffic Psychology and Behaviour* 51 (2017), 24–37. doi:10.1016/j.trf.2017.08.011

[19] Mouadh Guesmi, Mohamed Amine Chatti, Shoeb Joarder, Qurat Ul Ain, Rawaa Alatrash, Clara Siepmann, and Tannaz Vahidi. 2024. Interactive Explanation with Varying Level of Details in an Explainable Scientific Literature Recommender System. *International Journal of Human–Computer Interaction* 40, 22 (2024), 7248–7269. doi:10.1080/10447318.2023.2262797

[20] Jens Hainmueller, Dominik Hangartner, and Teppei Yamamoto. 2015. Validating Vignette and Conjoint Survey Experiments Against Real-World Behavior. *Proceedings of the National Academy of Sciences* 112, 8 (2015), 2395–2400. doi:10.1073/pnas.1416587112

[21] Judith A. Hall. 1980. Voice Tone and Persuasion. *Journal of Personality and Social Psychology* 38, 6 (1980), 924–934. doi:10.1037/0022-3514.38.6.924

[22] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. 2000. Explaining Collaborative Filtering Recommendations. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work* (Philadelphia, PA, USA) *(CSCW '00)*. Association for Computing Machinery, New York, NY, USA, 241–250. doi:10.1145/358916.358995

[23] Daniel Herzog, Hesham Massoud, and Wolfgang Wörndl. 2017. RouteMe: A Mobile Recommender System for Personalized, Multi-Modal Route Planning. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization* (Bratislava, Slovakia) *(UMAP '17)*. Association for Computing Machinery, New York, NY, USA, 67–75. doi:10.1145/3079628.3079680

[24] Jennifer Hettema, Julie Steele, and William R Miller. 2005. Motivational Interviewing. *Annual Review of Clinical Psychology* 1, 1 (2005), 91–111. doi:10.1146/annurev.clinpsy.1.102803.143833

[25] Åse Jevinger and Jan A Persson. 2019. Potentials of Context-Aware Travel Support During Unplanned Public Transport Disturbances. *Sustainability* 11, 6 (2019), 1649. doi:10.3390/su11061649

[26] Shalini Kapoor. 2025. Explainable and Context-Aware Graph Neural Networks for Dynamic Electric Vehicle Route Optimization to Optimal Charging Station. *Expert Systems with Applications* 283, Article 127331 (2025), 22 pages. doi:10.1016/j.eswa.2025.127331

[27] Ahyeon Kim, Minha Cho, Jungyong Ahn, and Yongjun Sung. 2019. Effects of Gender and Relationship Type on the Response to Artificial Intelligence. *Cyberpsychology, Behavior, and Social Networking* 22, 4 (2019), 249–253. doi:10.1089/cyber.2018.0581

[28] Seo Young Kim, Bernd H Schmitt, and Nadia M Thalmann. 2019. Eliza in the Uncanny Valley: Anthropomorphizing Consumer Robots Increases their Perceived Warmth but Decreases Liking. *Marketing Letters* 30, 1 (2019), 1–12. doi:10.1007/s11002-019-09485-9

[29] Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. 2019. Personalized Explanations for Hybrid Recommender Systems. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, CA, USA) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 379–390. doi:10.1145/3301275.3302306

[30] Robert M Krauss and Susan R Fussell. 1996. Social Psychological Models of Interpersonal Communication. In *Social Psychology: Handbook of Basic Principles*. The Guilford Press, New York, NY, USA, 655–701.

[31] Neal Lathia and Licia Capra. 2011. Mining Mobility Data to Minimise Travellers' Spending on Public Transport. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Diego, CA, USA) *(KDD '11)*. Association for Computing Machinery, New York, NY, USA, 1181–1189. doi:10.1145/2020408.2020590

[32] Rui Ma, Ning Liu, Jingsong Yuan, Huafeng Yang, and Jiandong Zhang. 2022. CAEN: A Hierarchically Attentive Evolution Network for Item-Attribute-Change-Aware Recommendation in the Growing E-commerce Environment. In *Proceedings of the 16th ACM Conference on Recommender Systems* (Seattle, WA, USA) *(RecSys '22)*. Association for Computing Machinery, New York, NY, USA, 278–287. doi:10.1145/3523227.3546773

[33] Eric B Meiners and Vernon D Miller. 2004. The Effect of Formality and Relational Tone on Supervisor/Subordinate Negotiation Episodes. *Western Journal of Communication* 68, 3 (2004), 302–321. doi:10.1080/10570310409374803

[34] John C Meyer. 2000. Humor as a Double-Edged Sword: Four Functions of Humor in Communication. *Communication Theory* 10, 3 (2000), 310–331. doi:10.1111/j.1468-2885.2000.tb00194.x

[35] Shiwali Mohan, Matthew Klenk, and Victoria Bellotti. 2019. Exploring How to Personalize Travel Mode Recommendations For Urban Transportation. In *Intelligent User Interfaces Workshops (CEUR Workshop Proceedings, Vol. 2327)*. CEUR-WS.org, Aachen, Germany, 6 pages. https://ceur-ws.org/Vol-2327/IUI19WS-HUMANIZE-2.pdf

[36] Kate G Niederhoffer and James W Pennebaker. 2002. Linguistic Style Matching in Social Interaction. *Journal of Language and Social Psychology* 21, 4 (2002), 337–360. doi:10.1177/026192702237953

[37] Ingrid Nunes and Dietmar Jannach. 2017. A Systematic Review and Taxonomy of Explanations in Decision Support and Recommender Systems. *User Modeling and User-Adapted Interaction* 27, 3 (2017), 393–444. doi:10.1007/s11257-017-9195-0

[38] Ayano Okoso, Keisuke Otaki, Satoshi Koide, and Yukino Baba. 2024. Toward Tone-Aware Explanations in Recommender Systems. In *Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization* (Cagliari, Italy) *(UMAP '24)*. Association for Computing Machinery, New York, NY, USA, 261–266. doi:10.1145/3627043.3659572

[39] Ayano Okoso, Keisuke Otaki, Satoshi Koide, and Yukino Baba. 2025. Impact of Tone-Aware Explanations in Recommender Systems. *ACM Transactions on Recommender Systems* 3, 4, Article 55 (April 2025), 34 pages. doi:10.1145/3718101

[40] Ayano Okoso, Mingzhe Yang, and Yukino Baba. 2025. Do Expressions Change Decisions? Exploring the Impact of AI's Explanation Tone on Decision-Making. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 824, 22 pages. doi:10.1145/3706598.3713744

[41] Jeroen Ooge, Leen Dereu, and Katrien Verbert. 2023. Steering Recommendations and Visualising Its Impact: Effects on Adolescents' Trust in E-Learning Platforms. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 156–170. doi:10.1145/3581641.3584046

[42] Maria Panou. 2012. Personalized Application for Multimodal Route Guidance for Travellers. *European Transport Research Review* 4, 1 (2012), 19–26. doi:10.1007/s12544-011-0067-0

[43] Sabid Bin Habib Pias, Ran Huang, Donald S. Williamson, Minjeong Kim, and Apu Kapadia. 2024. The Impact of Perceived Tone, Age, and Gender on Voice Assistant Persuasiveness in the Context of Product Recommendations. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 20, 15 pages. doi:10.1145/3640794.3665545

[44] Justus J Randolph. 2005. Free-Marginal Multirater Kappa (multirater K[free]): An Alternative to Fleiss' Fixed-Marginal Multirater Kappa. Online submission, presented at the Joensuu Learning and Instruction Symposium, Joensuu, Finland. https://eric.ed.gov/?id=ED490661 ERIC Document No. ED490661.

[45] Juan A Recio-García, Belén Díaz-Agudo, L Jose, and Marta Caro-Martínez. 2019. SAFEWAY: An Explainable Context-Aware Recommender System for Safe Routes. In *ICCBR Workshops (CEUR Workshop Proceedings, Vol. 2567)*. CEUR-WS.org, Aachen, Germany, 121–132. https://ceur-ws.org/Vol-2567/paper11.pdf

[46] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2022. Recommender Systems: Techniques, Applications, and Challenges. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer, New York, NY, USA, Chapter 1, 1–35. doi:10.1007/978-1-0716-2197-4_1

[47] Isabella Saccardi and Judith Masthoff. 2025. Adapting Emotional Support in Teams: Productivity, Emotional Stability, and Conscientiousness. *Frontiers in Artificial Intelligence* 8, Article 1449176 (2025), 18 pages. doi:10.3389/frai.2025.1449176

[48] Aaron Schild, Sreenivas Gollapudi, Anupam Gupta, Kostas Kollias, and Ali Sinop. 2025. Why is My Route Different Today? An Algorithm for Explaining Route Selection. In *2025 Proceedings of the Conference on Applied and Computational Discrete Algorithms* (Montreal, Quebec, Canada) *(ACDA 2025)*. SIAM, Philadelphia, PA, USA, 251–263. doi:10.1137/1.9781611978759.19

[49] Alistair Sutcliffe. 2017. Designing User Interfaces in Emotionally-Sensitive Applications. In *Human-Computer Interaction – INTERACT 2017*, Regina Bernhaupt, Girish Dalvi, Anirudha Joshi, Devanuj K. Balkrishan, Jacki O'Neill, and Marco Winckler (Eds.). Springer International Publishing, Cham, Switzerland, 404–422. doi:10.1007/978-3-319-67687-6_27

[50] Barbara G Tabachnick, Linda S Fidell, and Jodie B Ullman. 2007. *Using Multivariate Statistics* (5th ed.). Pearson, Boston, MA, USA.

[51] Nava Tintarev and Judith Masthoff. 2012. Evaluating the Effectiveness of Explanations for Recommender Systems: Methodological Issues and Empirical Studies on the Impact of Personalization. *User Modeling and User-Adapted Interaction* 22, 4 (2012), 399–439. doi:10.1007/s11257-011-9117-5

[52] Nava Tintarev and Judith Masthoff. 2022. Beyond Explaining Single Item Recommendations. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer, New York, NY, USA, Chapter 19, 711–756. doi:10.1007/978-1-0716-2197-4_19

[53] Marloes Vredenborg, Anouk van Kasteren, and Judith Masthoff. 2025. Personalization in Public Transport Passenger Information Systems: A Systematic Review and Framework. *Comput. Surveys* 57, 9, Article 226 (2025), 38 pages. doi:10.1145/3721478

[54] Yanyun Wang, Weizi Liu, and Mike Yao. 2025. Which Recommendation System Do You Trust the Most? Exploring the Impact of Perceived Anthropomorphism on Recommendation System Trust, Choice Confidence, and Information Disclosure. *New Media & Society* 27, 6 (2025), 3264–3292. doi:10.1177/14614448231223517

[55] James Ward, Jamie Sands, and Graham Johnson. 2007. The Effect of 'Tone of Dialogue' on Users' Interactions With Self-Service. In *Contemporary Ergonomics 2007: Proceedings of the International Conference on Contemporary Ergonomics* (Nottingham, UK) *(CE2007)*, Philip D. Bust (Ed.). Taylor & Francis, London, UK, 285–290. doi:10.1201/9781315106595

[56] Adam Waytz, Joy Heafner, and Nicholas Epley. 2014. The Mind in the Machine: Anthropomorphism Increases Trust in an Autonomous Vehicle. *Journal of Experimental Social Psychology* 52 (2014), 113–117. doi:10.1016/j.jesp.2014.01.005

[57] Shiming Zhang, Zhipeng Luo, Li Yang, Fei Teng, and Tianrui Li. 2024. A Survey of Route Recommendations: Methods, Applications, and Opportunities. *Information Fusion* 108 (2024), 102413. doi:10.1016/j.inffus.2024.102413