

MASTER'S THESIS

Visual attention and loneliness judgments: An eye-tracking study

Braah Baryhe

Faculty of Mechanical Engineering



DELFT UNIVERSITY OF TECHNOLOGY

MASTER'S THESIS

Visual attention and loneliness judgments: An eye-tracking study

By

B. Baryhe

In partial fulfilment of the requirements for the degree of:

Master of Science
in Biomedical Engineering
Track Medical Devices

at the Delft University of Technology,
to be defended publicly on Friday August 28, 2026 at 10:45 AM.

Supervisors: Dr. D. Dodou
J. Pfeifer, MSc

Thesis committee:	Dr. D. Dodou,	TU Delft
	Dr. A. Zgonnikov,	TU Delft
	Dr. ir. Y. B. Eisma,	TU Delft

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.

Contents

Abstract	4
1. Introduction	4
2. Study rationale, design, and hypotheses	7
2.1 Study aim	7
2.2 Single-person scene manipulations	7
2.3 Social-context manipulations	8
2.4 Hypotheses	9
3. Methods	10
3.1 Participants	10
3.2 Image generation and selection	10
3.3 Apparatus and software	11
3.4 Experiment procedure	12
3.5 Data processing	12
3.6 Dependent variables	13
3.7 Independent variables	13
3.8 Data analysis	14
4. Results	15
4.1 Descriptive analysis	15
4.2 Statistical analysis	18
4.3 Additional observations	20
5. Discussion	21
6. Limitations and future work	23
7. Conclusion	24
References	25
Appendix A	29
Appendix B	33
Appendix C	35

Abstract

Loneliness research has mainly focused on an inside-out perspective, examining how lonely individuals process social information. Less is known about the outside-in perspective: how people judge whether someone else appears lonely and which visual cues guide that judgment. Because loneliness is not always outwardly apparent, understanding how judgments of perceived loneliness are formed may help explain when loneliness becomes socially visible, when it remains overlooked, and when it is misattributed. This study examines how elements of a visual scene influence perceived loneliness and attention toward the person being judged. Participants ($N = 54$) viewed 102 AI-generated photorealistic images (six everyday scenes $\times$ 17 versions) while their eye movements were recorded. In each scene, a baseline image showed the foreground person alone, with a neutral expression in a neutral setting. Manipulated versions of each baseline scene changed the person's facial expression, the weather, the presence of objects, the background density, or the surrounding social situation. In the social scenes, further versions changed whether the foreground person remained neutral or reacted to the social situation. Each image was presented for 3 seconds, after which participants rated how lonely the foreground person appeared. Perceived loneliness was quantified as the rating shift from the matched baseline (absolute shift for magnitude comparisons); visual attention was quantified as dwell time percentage (DTP) on the foreground person's face. Among scenes depicting a single person, the facial expression manipulation produced the largest mean absolute rating shift and the highest face DTP of the four manipulation types. Changing the weather also produced relatively large mean absolute rating shifts, though these were not consistent across the weather variations, and reduced face DTP relative to baseline. The presence of objects and cluttered backgrounds mainly reduced face DTP without corresponding changes in perceived loneliness. In scenes depicting multiple people, the foreground person's reaction to the surrounding social situation was associated with a stronger influence on both loneliness judgments and face DTP than the type of social situation itself: reaction to the social situation produced larger mean absolute rating shifts and lower face DTP than remaining neutral. Overall, manipulations that produced the largest changes in loneliness judgments did not consistently produce the highest face DTP. This suggests that perceived loneliness was influenced by facial, environmental, and social information rather than attention on the face alone. However, how observers combined these cues was not directly examined here.

Keywords Perceived loneliness, Eye-tracking, Visual attention, Social context, Visual scene cues, AI-generated stimuli

1. Introduction

Loneliness, defined as the perceived gap between the social connection an individual needs or desires and the actual quality or quantity of interpersonal relationships (Perlman & Peplau, 1981), affects both quality of life and overall health, creating a significant challenge for public health (Rokach & Berman, 2025). Research has shown that loneliness is associated with adverse cardiovascular and cognitive outcomes, dementia and Alzheimer's disease, stroke, and mortality (Mushtaq et al., 2014). Beyond health consequences, loneliness frequently co-occurs with depression, generalized anxiety, and social anxiety with which it shares overlapping cognitive mechanisms such as low self-esteem and heightened sensitivity to social pain (Danneel et al., 2019; McDowell et al., 2021).

The cognitive model of loneliness proposed by Cacioppo and Hawkley (2009) and later extended by Qualter et al. (2015) suggests that loneliness activates cognitive processes that create biases in how lonely individuals attend to, interpret, and remember social information, particularly negative or socially threatening stimuli (Hawkley & Cacioppo, 2010; Spithoven et al., 2017). While this line of research takes an inside-out perspective, examining how loneliness shapes the lonely individual's perception of their surroundings and others, an outside-in perspective examines how observers form judgments of perceived loneliness from visual information and which visual cues within a scene they rely on when making these judgments. Such visual cues may be particularly important because loneliness is not always outwardly apparent, and observers may draw on the available cues when judging whether someone appears lonely. Understanding how judgments of perceived loneliness are formed may help explain when loneliness becomes socially visible, when it remains overlooked, and when it is misattributed. These judgments may, in turn, have implications for whether a person is noticed, approached, supported, or avoided.

Research has shown that social judgments are formed through the integration of multiple visual and situational cues in a scene (Kret & De Gelder, 2010; Wieser & Brosch, 2012). Studies have shown that faces are a primary source of information when judging another person's emotional and social state and are prioritized in visual attention (Birmingham et al., 2008; Dalmaso et al., 2025; Jaeger et al., 2022). Facial expressions communicate whether the emotional state appears positive or negative (valence) and how strongly that emotional expression appears (intensity) in a person (Akram & Stevenson, 2023). It has been shown that observers can generally identify basic emotions (e.g., happiness, anger, sadness, and fear) from the face with relatively high accuracy (Krumhuber et al., 2023). Studies have also shown that facial expressions influence social judgment, where observers were found to rely on facial expression when judging sociability-related traits, such as warmth and friendliness (Jaeger et al., 2022). At the same time, faces are not judged in isolation. Observers integrate facial expressions with the surrounding context to form an impression of the scene's emotional state (Wieser & Brosch, 2012). One example of such a contextual cue is weather. It has been shown to act as a background "atmosphere" that influences internal emotional state and the perception of the environment (Mei et al., 2023). For example, sunny weather is generally associated with a more positive atmosphere, and a rainy or gloomy one may make negative emotional states more noticeable. However, the influence of weather on internal emotional state has been found to be highly individual (Klimstra et al., 2011).

Beyond broader scene properties, localized objects can influence how visual attention is distributed within a scene. Objects may attract gaze because they are visually salient and stand out from surrounding elements (Schubö, 2009). They may also receive increased visual attention when they are semantically inconsistent with the scene, meaning that they do not match what observers expect to find in that setting (Pedziwiatr et al., 2022). As visually salient or semantically inconsistent elements, objects may therefore compete with other scene information for visual attention. Beyond individual objects, the overall density of a scene's background may also influence how it is visually explored. In photographs of real-world scenes, more visual clutter was associated with slower search, more missed targets, and longer average fixation durations. Viewers' first fixations tended to be directed towards more cluttered regions, although this tendency was not observed for later fixations (Henderson et al., 2009). More broadly, viewers can use learned knowledge about where objects are normally located in familiar types of scenes to direct gaze toward likely task-relevant locations (Henderson, 2017). These findings suggest that both the amount of visual information and the expected spatial arrangement of objects can influence how gaze is distributed within a scene. Although background structure and objects may influence how a scene is viewed, social information, especially faces, is generally prioritized even in the presence of other visually salient regions (Martinez-Cedillo & Foulsham, 2024).

Another element that has been studied is the presence of other people. Adding people alters the properties of a scene in a more complex way than the cues discussed so far because it introduces more social agents, and therefore more faces, behaviors, and contextual information. Research has shown that the presence of people in a scene attracts visual attention, as humans are evolutionarily inclined to prioritize social information such as faces and people (Birmingham et al., 2008; Dalmaso et al., 2025; Jaeger et al., 2022). Regarding perception, it has been found that a person's emotional expression was recognized more accurately when the surrounding social scene expressed a congruent emotion, for example a happy body was better recognized among people dancing at a party (Kret & De Gelder, 2010). Consistent with this, when an individual's facial expression matched the emotional context established by a preceding scenario, observers rated the individual as more likable and trustworthy (Chanes et al., 2018). This suggests that congruency between an individual's emotional expression, whether conveyed through body or face, and the surrounding social context can shape both how that expression is recognized and how the individual is judged.

Taken together, these findings suggest that several scene elements may contribute to both the perception of a person's emotional and social state and the way gaze is distributed across a scene, although not necessarily to the same extent or in the same direction for each element. If attention and judgment are related in this way, a method capable of measuring where observers look is needed; eye-tracking provides such a measure. This method records gaze position at high temporal resolution, measuring eye movements from which visual attention can be inferred, and thereby enables examination of how observers allocate attention across a scene (Holmqvist et al., 2011; Skaramagkas et al., 2023). Key attentional measures captured by eye-tracking

include early orientation (time to first fixation, location of first fixation) and sustained attention (fixation duration, fixation count, and dwell time) (Holmqvist et al., 2011). However, it is important to note that while eye-tracking objectively captures observable behaviors, the interpretation of these patterns as manifestations of underlying cognitive or affective processes requires contextual frameworks derived from self-report or theoretical models. While gaze and attention are tightly linked, they are not identical; a person can fixate directly on a target but fail to consciously process it (Mack & Rock, 1998).

Interpreting eye-tracking data therefore requires a theoretical model linking attention to broader cognitive processes. The human information processing framework describes how information from the senses is filtered by attention before being perceived, interpreted, and translated into a response (Wickens et al., 2013). This framework is general rather than specific to social perception and does not itself specify how social judgments such as perceived loneliness are formed. However, it offers a broad view in which attention helps select which visual cues in a scene are processed and contribute to the observer's perception. This does not imply that visual cues influence attention and judgment in the same way. Some eye-tracking studies in decision-making provide an analogy for a possible relationship between attention allocation and later judgments. Research in decision-making has shown that a longer dwell time on a specific option was associated with a greater likelihood of it being selected later (Pärnamets et al., 2016; Ting & Gluth, 2024). In another study using a pairwise AI image preference evaluation task, dwell time was higher for images that were later chosen than for images that were not chosen (Papadopoulos et al., 2026). In the same study, a machine-learning model was used to identify predictors of choice, and dwell time showed predictive accuracy above chance. This suggests that dwell time may function as a gaze-based predictor of later image choice. Although these studies concern preferential choice rather than social judgment, they demonstrate that gaze measures can provide information about how visual information is sampled during evaluative tasks. This provides a rationale for using gaze measures when examining how observers make judgments from visual stimuli.

As outlined previously, research has shown that social judgments are shaped by multiple visual cues. However, limited research has systematically isolated and compared these different element types within the same experimental framework to determine which of them influence perceived loneliness and whether they also redirect visual attention. A key challenge in conducting such a study is how to isolate individual scene elements while keeping all other aspects constant. AI generation of stimuli provides researchers with a tool to attempt the controlled manipulation of specific features (e.g., facial expressions) while keeping other factors (e.g., posture, clothing, lighting) as constant as possible. This can support greater experimental control such as the systematic isolation of key social cues (Ju et al., 2025; Renier et al., 2024; van Berlo et al., 2024), although it may also introduce artifacts, experimenter selection effects, and model dependent differences that may limit reproducibility. Despite the feasibility of combining eye-tracking and AI-generated media to study how visual cues influence gaze and perception, the application of this approach to examine loneliness perception remains largely unexplored.

The present study addresses this gap by using AI-generated stimuli to systematically manipulate six categories of visual cues: facial expression, weather, objects, background density, social context, and adaptation, defined here as whether the depicted individual's expression and behavior were adjusted to the surrounding social context. The manipulations were generated beforehand within controlled, standardized scenes. Participants then viewed these images in an eye-tracking task, during which loneliness ratings and eye-tracking data were recorded. Specifically, this study asks:

“How do visual scene manipulations influence observers’ judgments of loneliness and gaze toward the depicted person’s face?”

2. Study rationale, design, and hypotheses

2.1 Study aim

This study investigates which visual scene elements influence perceived loneliness of a depicted foreground person, and how these same elements affect visual attention toward that person's face, and whether patterns of judgment and gaze align. Six categories of visual cues were manipulated within AI-generated photorealistic images containing a foreground person, that is facial expression, weather, objects, background density, social context, and the foreground person's adaptation to that social context. The manipulations were applied, while all other aspects of the image were held as constant as possible. Each manipulated image was paired with a matched baseline image depicting the foreground person alone in a neutral setting with a neutral expression (see Section 3.2). Loneliness ratings of the foreground person were collected to capture the subjective perception of loneliness, and the dwell time percentage on the foreground person's face (hereafter face DTP), obtained via eye-tracking, was used to examine how the manipulations relate to visual attention toward the face. The manipulations were organized into two groups, reflecting a structural difference in what they change about the scene: single-person scene manipulations and social-context manipulations.

2.2 Single-person scene manipulations

The first group included four manipulation types applied to scenes depicting the foreground person alone: facial expression, weather, object, and background (see Table 1). These four types differed along two properties that motivated their selection: whether the manipulation was expected to change the social-emotional meaning of the scene and therefore perceived loneliness, and whether it introduced a visually prominent element that could compete with the foreground person's face for attention.

Facial expressions were included because they are a main source of information when observers judge another person's emotional and social state, including perceived sociability-related traits (Akram & Stevenson, 2023; Dalmaso et al., 2025; Jaeger et al., 2022). The manipulation was operationalized using four levels: a happy expression, a depressed expression, an expression suggesting social connectedness, and an expression suggesting social disconnectedness. Within each pair, the two levels were intended to produce changes of comparable magnitude but in opposite directions (happy & depressed, socially connected & socially disconnected). Because the manipulation directly altered the face of the person being judged, it was expected to directly influence perceived loneliness and visual attention in these scenes. Specifically, we predicted that depressed and socially disconnected faces would increase perceived loneliness, while happy and socially connected faces would decrease it.

Weather was included because observers integrate the surrounding environmental context with facial cues when forming social impressions (Wieser & Brosch, 2012). Weather has been found to influence internal emotional state and the perception of the environment (Mei et al., 2023). The manipulation was operationalized using two levels: holiday weather, intended to convey a sunny vacation-like atmosphere, and depressing weather, intended to convey a gloomy, rainy, and wet atmosphere. As a diffused manipulation, weather was not intended to introduce a single localized element; therefore, it was not expected to compete with the face for visual attention but was instead expected to subtly change the general atmosphere of the scene.

Objects were included as localized elements positioned outside the face that could attract gaze because they were visually salient or semantically inconsistent with the surrounding scene (Pedziwiatr et al., 2022; Schubö, 2009). The manipulation was operationalized using two levels: the "social" object level included an object intended to subtly imply the presence of another person (e.g., a left-behind bag), a design choice inspired by an approach used in photography (Bell, 2014). The "non-social" object level (e.g., a boulder) included an inanimate object that was semantically inconsistent with the scene. In both levels, the bottom-up salience of the object was increased by specifying bright and contrasting colors in the prompt. Because the object did not show an actual social interaction and only subtly implied it in the peripheral areas of the scene, the object manipulation was not expected to strongly alter the social meaning of the scene, nor produce a large or consistent shift in perceived loneliness. The main expectation, however, was for the object to compete with other scene information for visual attention and potentially reduce face directed gaze,

although the face was still expected to remain an important focus of attention and source of social information (Martinez-Cedillo & Foulsham, 2024).

Background was included as a diffuse, peripheral manipulation intended to alter the amount of visual information in the scene without introducing a single localized or strongly meaningful cue. The manipulation was operationalized using two levels. The cluttered level added diffuse additional clutter across the background, whereas the sparse level created a more vacant looking environment by removing background elements. Although background clutter can generate bottom-up attentional competition (Henderson et al., 2009), its influence was expected to be limited by top-down task relevance (Henderson, 2017). Because the task required participants to judge the foreground person, background information was expected to be less task-relevant than person-related cues. Background was therefore treated as a near-control manipulation, with its effects expected to be attenuated relative to more localized and socially meaningful cues, particularly the face. Accordingly, little or no change in either loneliness judgments or face directed attention was predicted relative to baseline (Martinez-Cedillo & Foulsham, 2024).

These four conditions (facial expression, weather, object, and background) were analyzed as a single four-level within-subjects factor (manipulation type), and the reasoning above motivates the expected ordering of their effects (see Table 1).

Table 1: Single-person manipulations

Manipulation type	Levels	Predicted influence on the scene's social-emotional meaning	Localized or diffused
Facial expression	Depressed, happy, socially connected, socially disconnected	Direct	Localized
Weather	Weather holiday, weather depressing	Indirect, through the scene atmosphere	Diffused
Object	Object social, object non-social	Small or subtle	Localized
Background	Background cluttered, background sparse	Little or no influence; designed as a near-control	Diffused

Note: Hereafter, manipulation type refers to the broader cue category (facial expression, weather, object, or background), whereas condition level refers to a specific variant within that manipulation type (e.g., depressing weather).

2.3 Social-context manipulations

The second group was treated separately because adding other people changes the scene in a more complex way than the single-person manipulations do. Adding people changes the relationship between the foreground person and others, the spatial composition of the scene, the total amount of social information, and the number of potential targets for attention.

The presence and behavior of other people in a scene were included because adding people introduces additional social agents, faces, behaviors, and contextual information that can shape both visual attention and social-emotional judgments (Kret & De Gelder, 2010; Dalmasso et al., 2025). The manipulation was operationalized using three social-context levels: a dyad involving one additional person, a group of friends, and a group excluding the foreground person, with the latter two involving more than one additional person. Because these three levels differ in how many additional people are present, they were also expected to differ in how strongly the added faces competed for gaze toward the foreground person's face, with scenes containing more people expected to draw relatively more attention away. In terms of perception, specifically exclusion has been connected to loneliness in previous research as both are linked through a shared underlying mechanism: a threat to the fundamental need to belong (Basset et al., 2025; DeWall et al., 2009; Spithoven et al., 2017). Exclusion scenes may, therefore, lead observers to view the foreground person as socially disconnected and are expected to produce the largest increase in loneliness ratings among the three context types.

Adaptation refers to whether the foreground person's expression and behavior matched the surrounding social situation. For each social-context level, the non-adapted version showed the foreground person

remaining neutral, whereas the adapted version showed a response that matched the social situation. In the dyad and friends scenes, this response involved positive social engagement, while in the exclusion scenes it involved a negative reaction to being excluded. This variant was included because prior work has shown that compatibility between a person's expression, bodily or facial, and their social context influences how observers judge that person (Chanes et al., 2018) and their emotional state (Kret & De Gelder, 2010). Adaptation was therefore expected to make the social meaning of the scene less ambiguous and support a more decisive judgment of the foreground person's loneliness. This was expected to produce a larger magnitude of loneliness rating shift relative to baseline.

Context and adaptation were organized in a fully crossed two-factor within-subjects design. The aim was to test whether changes in perceived loneliness and visual attention are driven primarily by context, adaptation, or by their interaction (see Table 2).

Table 2: Social-context manipulations

Context/Adaptation	Not adapted (foreground person neutral)	Adapted (foreground person reacts)
Dyad	Dyad present	Dyad present adapted
Friends	Friends present	Friends present adapted
Exclusion	Others excluding	Others excluding adapted

2.4 Hypotheses

H1: The magnitude of loneliness rating shift, relative to the matched baseline image, is expected to differ across the four manipulation types, following the ordering: facial expression > weather > {object $\approx$ background}, with object and background expected to produce the smallest shifts and not to differ reliably from each other.

H2a: Face DTP is expected to differ across the four manipulation types, following the ordering: facial expression > {weather $\approx$ background} > object, with weather and background not expected to differ reliably from each other.

H2b: Relative to baseline, face DTP is expected to be reduced by the object manipulation, which introduces a discrete competing element. The facial expression, weather, and background manipulations are not expected to differ significantly from baseline, since none of these introduces a discrete element outside of the face area.

H3: For the social-context scenes, adapted conditions are expected to produce larger absolute loneliness rating shifts than their non-adapted counterparts, and among the three context types, exclusion is expected to produce a larger shift than dyad and friends.

H4: Dyad is expected to produce the highest face DTP among the three social-context types.

Effects not specified in H1–H4, including the Context $\times$ Adaptation interactions and the effect of adaptation on face DTP, were examined exploratorily.

3. Methods

3.1 Participants

The sample consisted of 54 participants, with a mean age of 23.9 years ($SD = 2.9$, range = 18–31). The sample included 29 female participants (53.7%), 24 male participants (44.4%), and one participant who preferred not to report their gender. Participants represented 21 nationalities, with Dutch being the majority ($n = 24$), followed by Spanish ($n = 4$), mixed nationality ($n = 3$), and several other nationalities represented by one or two participants each. During the experiment, 4 participants wore glasses (7.4%), 11 wore contact lenses (20.4%), 10 usually wore visual correction but did not wear it during the experiment (18.5%), and 29 reported using no visual aids (53.7%). Trait loneliness was assessed using the UCLA 3-Item Loneliness Scale (Hughes et al., 2004). Each item was rated as “hardly ever” (1), “some of the time” (2), or “often” (3). The three responses were summed to produce a total score ranging from 3 to 9, with higher scores indicating greater trait loneliness. The mean total score was 5.28 ($SD = 1.28$, range = 3–9). State loneliness was measured using a single item asking participants how lonely they felt at the time of testing (adapted from Burgin et al., 2012). Responses were rated as “not at all” (1), “a little bit” (2), “somewhat” (3), “quite a bit” (4), or “very much” (5). Most participants reported feeling either “not at all” lonely ($n = 21$, 38.9%) or “a little bit” lonely ($n = 22$, 40.7%) at the time of testing. The research was approved by the Human Research Ethics Committee of Delft University of Technology (approval number 6950), and all participants provided their written informed consent.

3.2 Image generation and selection

Achieving control over individual scene elements while holding all other aspects in the image constant is difficult with conventional photography or even photo editing tools. To address this, the full stimulus image dataset was generated using an automated pipeline developed by De Winter (2026) for this project. The pipeline first created six everyday-setting baseline scenes using OpenAI’s GPT-Image-2, involving one person situated in the foreground. Three of those scenes were set outdoors (i.e., university outdoors, park, and train station), and the other three were set indoors (i.e., university indoors, cafe, and home living room). The scenes were then balanced for the sex of the foreground person (3 male, 3 female), with two male and one female in the outdoor scenes and one male and two female in the indoor scenes, and their position in the frame (3 left, 3 right). Each of the baseline images represented the foreground person alone in a neutral setting with a neutral facial expression. The baseline scenes were manipulated by applying condition level-specific edit instructions for each of the 16 manipulation levels (see Table 1 and Table 2), creating 17 images per scene (102 overall images; 6 scenes $\times$ 17 images). A key constraint was that all the other elements in the image, including the foreground person’s identity, clothing, and camera angle remained unchanged and as consistent as possible across condition levels within each scene. Weather and background manipulations also introduced changes to related visual features, such as lighting, season, color, and the arrangement of scene elements (see Figure 1a). The adaptation condition (see Table 2) followed a different editing route, where the edits were applied to the corresponding non-adapted image instead of the baseline, so that the added people remained identical and only the foreground person changed (see Figure 1b).

The code produced five candidate stimulus sets. The final set used in the study was then selected manually from these candidates, choosing the set with the fewest AI artifacts and hallucinations (see Figure 1).



Figure 1: Example stimulus scene set showing one baseline scene and its 16 manipulated versions across all condition levels. Panel (a) shows the single-person scene manipulations, comprising four manipulation types and their corresponding levels: facial expression, weather, object, and background. Panel (b) shows the social-context manipulations, comprising three context types, that is dyad, friends, and exclusion, with a non-adapted and adapted version of each. Both panels repeat the same baseline image, in which the foreground person appears alone in a neutral outdoor university setting. Across the stimulus set, the foreground person's identity, clothing, position in the frame, and camera angle were kept as consistent as possible, while only the condition level-specific element was changed. For the adapted social-context conditions, the edits were applied to the corresponding non-adapted image instead of the baseline image so that the added people remained identical and only the foreground person changed.

3.3 Apparatus and software

The experiment design and interface were programmed in the SR Research Experiment Builder (version 2.3.38). The experiment ran on a PC (64-bit Windows 7 Professional). Monocular eye movements from the right eye were recorded at 2000 Hz using the SR Research EyeLink 1000 Plus. Head position was stabilized with a chin-and-forehead rest mounted at the table edge. The monitor was positioned approximately 93–95 cm from the participants' eyes, with measured distances of 93 cm to the top and 95 cm to the bottom of the visible display area of the monitor. The horizontal distance between the eye-tracking camera lens and the participants' eyes was approximately 53 cm (see Figure 2). The stimuli were presented on a 24.5-inch BenQ monitor (XL2540-B) with a resolution of 1920×1080 pixels and a refresh rate of 120 Hz. The monitor occupied approximately 32.5° horizontally and 18.9° vertically of the participant's field of view.

The illuminance, measured with a Konica Minolta T-10MA at 95 cm from the screen, corresponding to the approximate location of the participant's eyes, was 201 lx when the sensor was pointed toward the ceiling and ranged from 113.7 to 150.3 lx when pointed toward the monitor across display intensity levels. Luminance, measured with a Konica Minolta LS-150 at 100 cm from the screen, was 28.70 cd/m^2 for a grey surface [RGB: 127, 127, 127], 1.040 cd/m^2 for a black surface [RGB: 0, 0, 0], and 158.1 cd/m^2 for a white surface [RGB: 255, 255, 255].

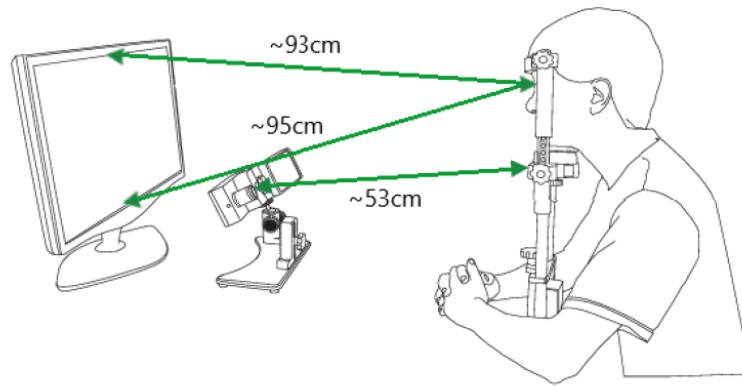


Figure 2: Schematic illustration of the eye-tracking physical setup. Participants were seated in front of the display monitor with their head stabilized using a chin and forehead rest. The monitor was positioned approximately 93–95 cm from the participant’s eyes, with a distance of approximately 93 cm to the top of the display and approximately 95 cm to the bottom of the display. The eye-tracking camera was positioned between the participant and the monitor, approximately 53 cm from the participant’s eyes.

3.4 Experiment procedure

Before starting the study, participants signed the consent form. They then completed the standard 9-point calibration and validation for the EyeLink 1000 Plus system. Afterward, they were presented with these instructions: “During the task, each image will be displayed for a few seconds. In between each image, a cross will be presented at the center of the screen. Please look at this. After each image, rate how lonely you think the foreground (closest to the front) person is. Consider all available information in the image: the person’s facial expression and body posture, the setting, and the presence and behavior of any other people. There is no right or wrong answer. Press any key to begin the task”.

Each trial began with a central fixation cross displayed for 700 ms, followed by a single image displayed for 3000 ms. The image measured 1505×1006 pixels spanning from (163, 39) to (1667, 1044) screen coordinates on a 1920×1080 resolution screen. After each image, participants rated how lonely the foreground person seemed using a visual analog scale (VAS) slider that ranged from “extremely lonely” to “not lonely at all”. The scale contained 11 equally spaced, unnumbered tick marks, and participants could position the slider anywhere within the scale rather than only on the tick marks. The instruction stated: “Consider all available information in the image. Rate how lonely you think the foreground person is”.

Eye movements were recorded during image presentation only, while ratings were recorded during the rating slider display phase. The experiment consisted of 102 trials, corresponding to 6 scenes $\times$ 17 condition levels. Each participant viewed the entire dataset in a fully randomized order, without any practice trials or scheduled breaks. Participants then completed a questionnaire (via Qualtrics) on a separate computer. The questionnaire included demographic and vision related questions, open ended exploratory questions about participants’ judgment strategies and perceptions of the stimuli, the three item UCLA Loneliness Scale measuring trait loneliness, and a closed-ended question assessing current loneliness. The full procedure took approximately 30 minutes, although the duration varied across participants.

3.5 Data processing

The onset and offset of each image viewing epoch were determined from the eye tracker’s event messages, which logged the exact time each image was shown on the screen. Image onset was defined as the time at which the stimulus image appeared on the screen, and image offset was defined as the time at which the following rating scale screen appeared. To check whether the image viewing epochs had been extracted correctly, each extracted epoch duration was compared with the expected 3000 ms image viewing period, allowing a ± 500 ms tolerance. Durations outside the 2500–3500 ms range were flagged but not excluded. No epochs fell outside this range. Gaze samples located more than 50 pixels outside the screen boundaries of the 1920×1080 display were classified as invalid and set to missing. Missing gaze periods, including blinks and tracking losses, were identified from gaps in either horizontal (x) or vertical (y) gaze signal. To account for the gradual eyelid closure and reopening, a 100 ms margin was added before and after each gap (Caffier et al., 2003).

Two levels of eye-tracking data quality control were applied. First, at the trial level, trials were excluded from further analyses when more than 50% of all samples within the image viewing epoch were marked as missing or invalid in the expanded missing data mask. Second, at the participant level, participants with more than 50% of invalid trials (i.e., more than 51 trials) were excluded entirely from the eye-tracking analyses. Therefore, it was possible for individual trials to be excluded while no participants were excluded at the participant level.

For the retained trials, linear interpolation was applied to both the horizontal (x) and vertical (y) gaze coordinates of the missing segments. Missing data points at the beginning or end of a trial were filled using the nearest valid value. The interpolated gaze coordinates were then smoothed using a moving median filter with a 100 ms window. In total, 34 trials from 13 participants were excluded at the trial level, and no participants were excluded at the participant level, leaving 99.38% of the total 5508 (54 participants $\times$ 102 trials per participant) trials for analysis.

For each scene two areas of interest (AOIs) were defined: the foreground person face, and the remaining image area. The foreground person face AOI was defined manually as a single rectangular bounding box per scene, specified as [$x_{upper\ left-corner}$, $y_{upper\ left-corner}$, $width$, $height$] in image space, and visually verified. Because some manipulations (e.g., the adaptation manipulation) altered the foreground person's pose, the face appeared in slightly different positions. To keep the AOI identical across conditions within a scene, the box was drawn to contain the face in all conditions. No fixed margin or padding was added; the box was fitted by hand to be as tight as possible while still enclosing the face in every condition. Therefore, the box bounds the face more loosely in some images than in others.

Across the six scenes, the face AOI occupied on average 4.27 % of the image (SD = 1.97%, range 1.43–7.49 %). For each stimulus image, these rectangular AOI definitions were converted into pixel-level AOI masks. In these masks, pixels inside the foreground face rectangle were labeled as the face AOI, while the remaining pixels within the displayed image were labeled as the rest of the image. The masks were then resized to the displayed image dimensions of 1505 $\times$ 1006 pixels and placed into screen coordinates. During the analysis, gaze samples' locations were assigned to AOIs by rounding the gaze position to the nearest pixel and matching it to the corresponding AOI mask. Gaze samples that fell on screen but outside the displayed image were not included in the face DTP analysis.

3.6 Dependent variables

3.6.1 Eye-tracking metrics

Face dwell time percentage (DTP): defined as the number of valid gaze samples (all x, y coordinates sampled, i.e., fixations and saccades) falling within the foreground person face AOI divided by the number of valid gaze samples falling within the image area, multiplied by 100%.

3.6.2 Behavioral

Rating shift: defined as the difference between the loneliness rating for a given condition level and the baseline rating for the same scene. Ratings were collected using a visual analogue scale with “extremely lonely” displayed at the left endpoint and “not lonely at all” at the right endpoint. The scale contained 11 equally spaced tick marks without numerical labels, and participants could position the slider anywhere along the scale. For analysis, slider positions were coded from 0, corresponding to “not lonely at all,” to 100, corresponding to “extremely lonely,” so that higher scores indicated greater perceived loneliness. Responses were rounded to the nearest integer, with a numerical resolution of one scale point, equivalent to approximately 13.2 screen pixels.

3.7 Independent variables

3.7.1 Single-person scene manipulations

Manipulation type was a single within-subjects factor with four levels, each altering one element of a scene depicting the foreground person alone (see Table 1):

- (i) Facial expression: the facial expression of the foreground person, with four condition levels (depressed, happy, socially connected, socially disconnected).

- (ii) Weather: the weather and atmospheric conditions of the scene, with two condition levels (depressing, holiday).
- (iii) Object: a visually salient object introduced into the scene outside the face region, with two condition levels (social, non-social).
- (iv) Background: the density and spatial layout of the scene's background, with two condition levels (cluttered, sparse).

3.7.2 Social context manipulations

The social-context conditions comprised two within-subjects factors, fully crossed (see Table 2):

- (i) Context: the type of social context present in the scene, with three levels. In the dyad condition, one additional person was present besides the foreground person. In the friends and exclusion conditions, more than one additional person was present; in exclusion, the foreground person was depicted as being excluded by the others.
- (ii) Adaptation: the state of the foreground person in relation to the added social context, with two levels. In the non-adapted level, the foreground person's expression remained neutral regardless of the surrounding context. In the adapted level, the foreground person reacted accordingly to the social situation.

3.8 Data analysis

Descriptive statistics were calculated for loneliness ratings, rating shifts from baseline, and face DTP. For each participant, trial-level values were first averaged across the six scenes within each condition level, giving one value per participant for each condition. These participant-level values were then summarized across participants ($n = 54$) using the mean, standard deviation, and 95% confidence interval (CI). Confidence intervals were calculated across participants as:

$$CI = \text{mean} \pm t(0.975, n - 1) \times SD/\sqrt{n}$$

Equation 1: Calculation of the 95% confidence interval around the sample mean, where $t_{0.975, n-1}$ is the critical value from the t-distribution with $n-1$ degrees of freedom, SD is the sample standard deviation, and n is the number of participants.

Loneliness rating effects were calculated, for each participant and scene, as the difference between each manipulated condition-level rating and the baseline rating of that scene, which was then averaged across scenes. Signed and absolute loneliness rating shifts were used to describe the direction and magnitude of the rating change compared to the baseline, respectively, and were visualized using bar plots and forest-style plots. Temporal changes in face DTP were analyzed by calculating DTP across the full image viewing trial time (3000 ms). Only trials containing the complete first 3000 ms were included, and longer trials were truncated to 3000 ms; no trials were excluded on this basis, as all trials passed the criteria. Following the same aggregation as above, DTP values were averaged across scenes within each participant and then summarized across participants at each time sample. For visualization, the resulting time courses were downsampled to one point every 10 ms and plotted per condition level as line plots with 95% CIs. To complement the quantitative DTP analysis, descriptive gaze heatmaps were generated. For each stimulus image, filtered gaze coordinates (as described in section 3.5) were collected across participants and trials. The gaze coordinates were converted from screen to image coordinates and accumulated into a density map, which was divided by the total number of gaze samples for that image and then smoothed with a Gaussian kernel. The smoothed maps were normalized using a single maximum value calculated across all stimulus images, allowing the heatmap colors to follow a shared scale across images, and a gamma correction was applied to improve the visibility of lower density regions. The color-coded density maps were overlaid on the original stimulus image, with warmer colors indicating higher gaze density and cooler colors indicating lower gaze density.

Inferential analyses were conducted using repeated-measures ANOVAs on participant level condition means using face DTP and absolute rating shift as the dependent variables. Because condition levels could shift ratings in opposite directions, averaging signed shifts could produce a mean near zero. Absolute rating shift was therefore used to compare the overall magnitude of change across conditions. Before conducting

the repeated measures ANOVAs, the relevant assumptions were assessed (Muhammad, 2023). These included extreme outliers check using a ± 3 SD criterion, checking for normality of the dependent variables using the Lilliefors test, and sphericity where a within-subject factor had three or more levels. Four repeated-measures ANOVAs were conducted. Two examined the single-person manipulations, with manipulation type (facial expression, weather, object, background) as a single four-level within-subjects factor: Model A examined absolute rating shift and Model B examined face DTP. The remaining two examined the social-context manipulations, with context and adaptation entered as two crossed within-subjects factors: Model C examined absolute rating shift, and Model D examined face DTP. For each model, F-values, p-values, and partial eta squared (η^2p) were reported.

Follow-up analyses were conducted to interpret the repeated-measures ANOVA results. For the single-person scene manipulations (facial expression, weather, object, background), paired-sample t-tests compared the four manipulation types on absolute rating shift and face DTP. In addition, each individual condition level was tested against its matched baseline. For face DTP, paired-sample t-tests compared each level against the baseline condition. For rating, one-sample t-tests compared the signed rating shift against zero, since the signed rating shift is itself the difference from the matched baseline image. Bonferroni-corrected p-values were reported separately for the pairwise comparisons ($n = 6$) and the individual condition levels baseline comparisons ($n = 10$). For the social-context manipulations, Tukey–Kramer adjusted post-hoc pairwise comparisons were conducted for the three-level context factor in the absolute rating shift model and face DTP, comparing dyad, friends, and exclusion levels. Adaptation had only two levels and therefore did not require post-hoc pairwise comparisons beyond the main effect tested in the repeated-measures ANOVA. Each of the six social-context levels was additionally tested against its matched baseline, using one-sample t-tests against zero for signed rating shift and paired-sample t-tests for face DTP ($n = 6$).

Finally, two descriptive scatter plots were used to visualize the relationship between mean face DTP and mean rating shift from baseline. The first analysis summarized the 16 non-baseline condition levels using signed rating shifts, allowing both the magnitude and direction of the rating change to be visualized. Each point represented the mean signed rating shift and mean face DTP for one condition level. The second analysis was conducted at the broader condition level and included 10 points: the four single-person conditions and the six social-context conditions. Each point represented the mean absolute rating shift from baseline and mean face DTP. All analyses were conducted in MATLAB R2024a.

4. Results

4.1 Descriptive analysis

To describe how each condition level changed the foreground person’s perceived loneliness, we compared loneliness ratings for each level with the matching baseline scene (see Figure 3). In the single-person manipulations group, facial expression levels produced the largest shifts of any condition, in both directions (face depressed: +27.1; face happy: -23.4). Weather depressing also produced a substantial positive shift (+12.7) indicating higher loneliness rating than baseline while weather holiday remained relatively close to the baseline (-2.2). Object and background levels showed minimal shifts compared to baseline, although background showed some asymmetry, with background-sparse (+6.4) producing a larger shift than background-cluttered (-1.3).

In the social-context manipulation group, both exclusion levels showed positive shifts, indicating higher loneliness ratings than baseline. The adapted exclusion level showed the largest positive shift across all social-context levels (+34.5), compared with (+11.4) for the non-adapted exclusion level. In contrast, “dyad present adapted” showed the largest shift in the opposite direction (-30.1), indicating lower loneliness ratings than baseline, whereas the non-adapted “dyad present” level remained close to baseline (see Figure 3). A descriptive overview of the mean loneliness ratings across all 17 condition levels is provided in Appendix B, Figure B.1.

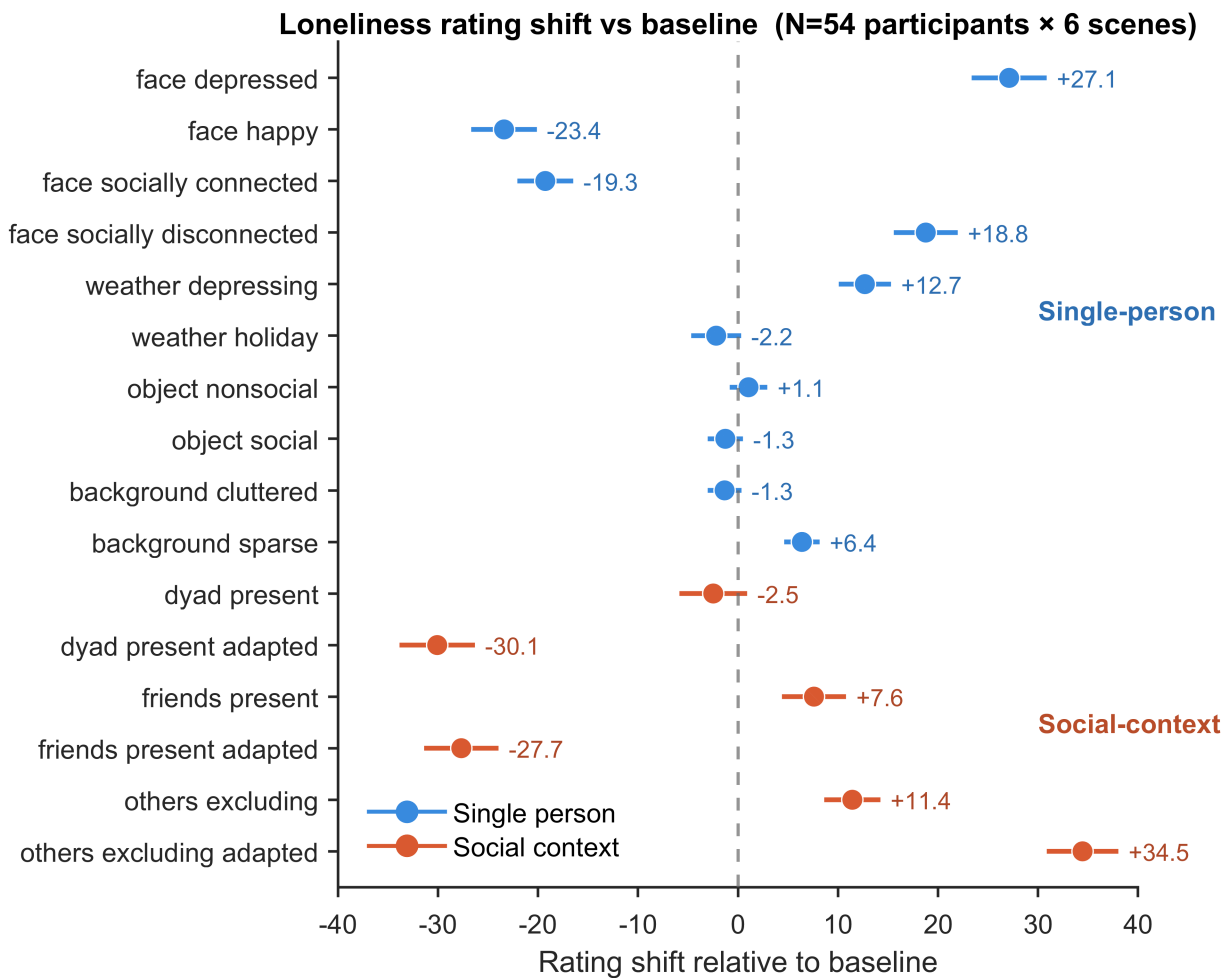


Figure 3: Loneliness rating shifts from matched baseline across condition levels. Points show the mean signed rating shift for each non-baseline condition level, calculated as the rating for that level minus the rating of the matched baseline scene. Rating shifts were averaged across scenes and then across the 54 participants. Positive values indicate higher loneliness ratings than baseline, and negative values indicate lower loneliness ratings than baseline. Blue points represent single-person manipulation levels, and red points represent social-context manipulation levels. Error bars indicate 95% confidence intervals, and the dashed vertical line marks zero shift, indicating no change from baseline.

For the single-person scene manipulations, face DTP showed a clear early peak shortly after image onset, approximately between 300 and 600 ms, followed by a gradual decrease across the remaining viewing period (see Figure 4a). This pattern was visible across all manipulation types, but the magnitude and later temporal pattern differed between conditions. The facial expression manipulation levels remained relatively close to baseline after the initial peak, suggesting that face directed attention was sustained across the viewing window. In contrast, the object manipulations showed a noticeable decrease, compared to baseline, in both the early peak magnitude and in the following part of the trial. This suggests that participants may have redirected attention away from the foreground person face toward other elements in the image, most likely the social/non-social object as suggested by the descriptive heatmaps (see Figure 6). Weather and background manipulations settled below baseline after the early peak, with the background levels showing asymmetry: the sparse background followed a trajectory closer to baseline, whereas the cluttered background showed a more noticeable reduction in face DTP.

For the social-context manipulations, all conditions showed lower face DTP than baseline across most of the viewing period (see Figure 4b). Overall, the adapted levels tended to show lower face DTP than their non-adapted counterparts during later parts of the trial, particularly in the friends and exclusion contexts. Although this pattern was not uniform across the full time course. Both dyad levels reached an earlier peak than the other social-context conditions, followed by a noticeable decrease in face DTP. Between approximately 500 and 1000 ms, the adapted dyad level showed higher DTP than the non-adapted level, reversing the general pattern, in which adapted conditions showed lower DTP than non-adapted ones. The descriptive heatmaps suggest that, in social-context scenes, participants looked at other people's faces in

addition to the foreground person's face (see Figure 6). This may suggest that the decrease in foreground face DTP reflects attention shift to the other socially relevant faces in the scene.

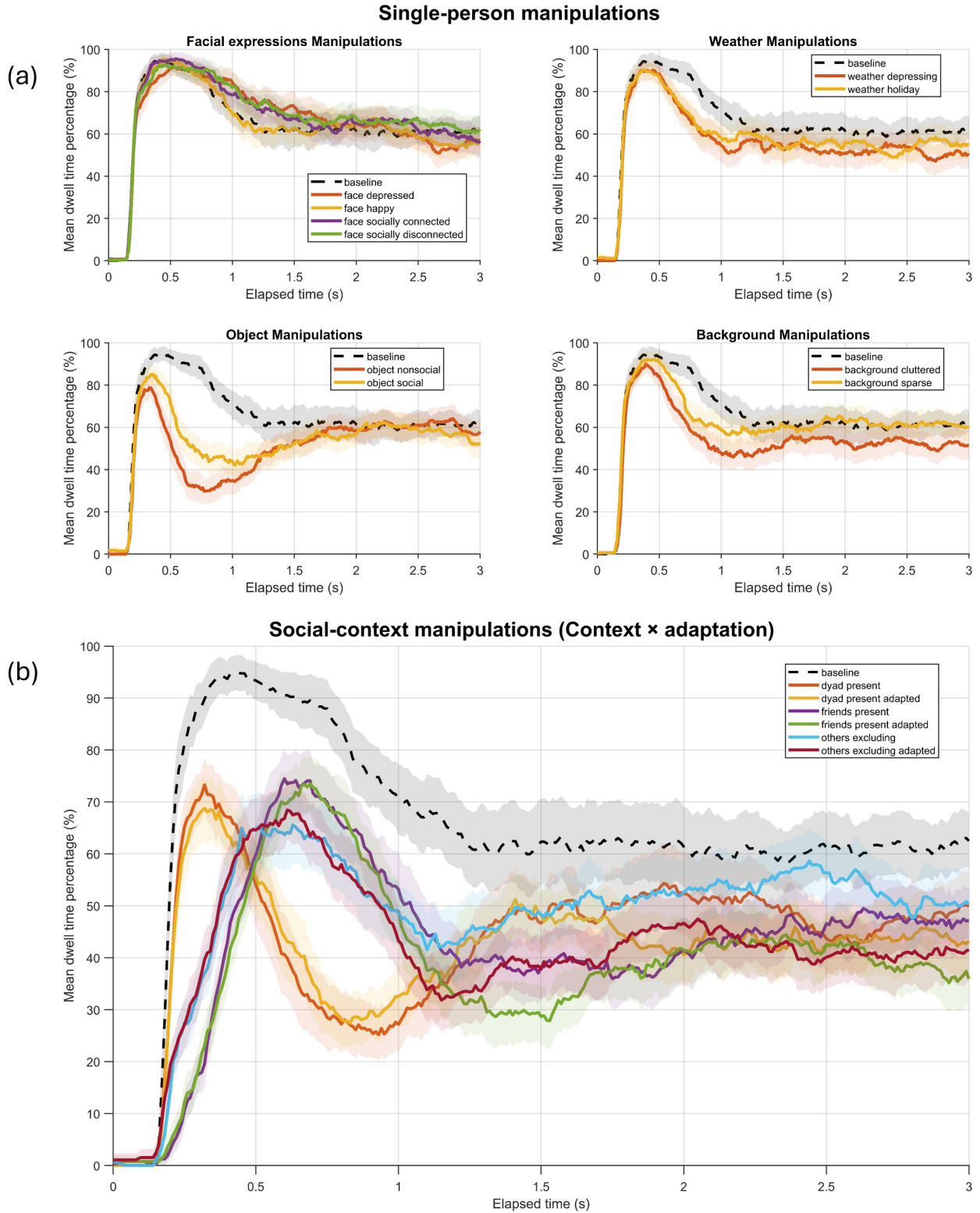


Figure 4: Time course of visual attention to the foreground face AOI across the 3 s image viewing period, with 95% confidence intervals. The graphs show the mean dwell time percentage of valid in-image gaze samples falling within the foreground person's face AOI over time, averaged across the six scenes within each participant and then across participants ($n = 54$). Shaded regions represent 95% confidence intervals. For visualization, the time courses were downsampled to 10 ms intervals. Panel (a) shows the single-person scene manipulations, that is facial expression, weather, object, and background manipulations. Panel (b) shows the social-context manipulations, that is dyad, friends, and exclusion conditions with non-adapted and adapted versions. Full DTP values for all condition levels are provided in Appendix A, Table A1.

Finally, descriptive scatter plots were used to visualize the connection between face DTP and signed loneliness rating shift, relative to matched baseline, across manipulations' condition levels (see Figure 5). The plot summarizes the 16 non-baseline condition levels belonging to single-person and social-context manipulations (see Table 1 and Table 2). The plot shows that, among the single-person condition levels, the facial-expression levels produced the largest shifts in loneliness ratings in both directions from baseline, while their face DTP values remained relatively high and close to the mean baseline face DTP. In contrast, object levels, background cluttered, and weather holiday stayed closer to baseline in both outcomes, with object social and background cluttered overlapping in the plot. Social-context manipulations mostly reside in the relatively low face DTP range, with the adapted levels moving the furthest in loneliness rating in either direction from baseline, and the non-adapted levels remaining closer to baseline than their counterparts.

An additional manipulation scatter plot is provided in Appendix B, Figure B.2.

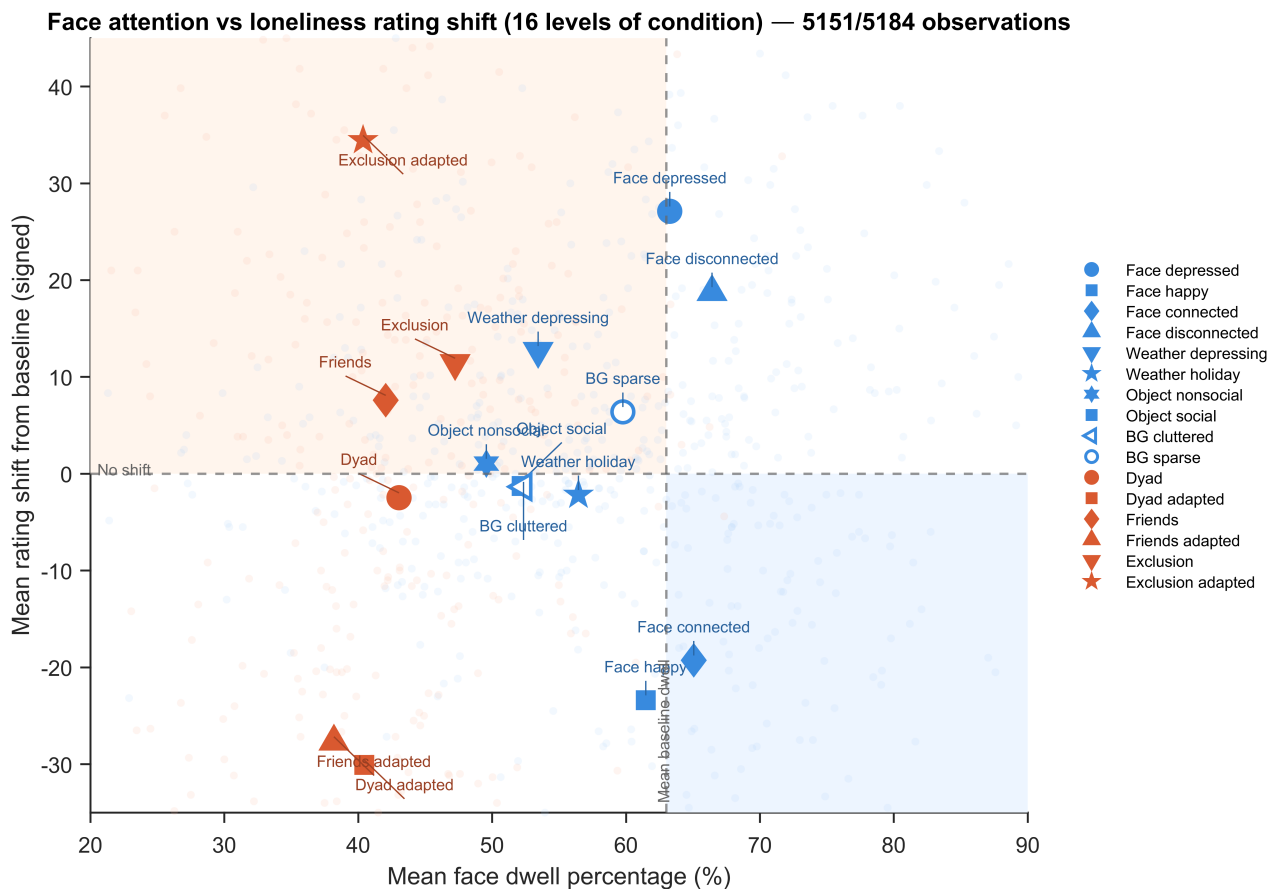


Figure 5: Descriptive comparison between face dwell time percentage and signed loneliness rating shift across condition levels. The scatter plot shows the mean face dwell time percentage (DTP) against mean signed rating shift from baseline across the 16 non-baseline condition levels. Large markers represent the group mean for each condition level, with single-person manipulation levels shown in blue and social-context manipulation levels shown in red. Faint background points show participant-level values: each point is one participant's mean face DTP and mean rating shift for one condition level (54 participants $\times$ 16 condition levels). The x-axis represents mean face DTP, and the y-axis represents the signed rating shift from the matched baseline scene. Positive values indicate higher loneliness ratings than baseline, whereas negative values indicate lower loneliness ratings than baseline. The vertical dashed line marks the mean face DTP across baseline scenes, and the horizontal dashed line marks zero shift, indicating no change from baseline in face DTP. A total of 5151 out of 5184 expected observations (16 non-baseline condition levels $\times$ 6 scenes $\times$ 54 participants) were included after data filtering.

4.2 Statistical analysis

ANOVA assumption checks showed only limited deviations from normality in a small number of cells. Four out of 20 model cells (10 experimental condition cells $\times$ 2 outcomes: absolute rating shift and face DTP) showed significant deviations. Five outliers beyond ± 3 SD were identified across the 20 model cells. All outliers were retained. Mauchly's test indicated a violation of sphericity for the single-person Model A (absolute rating shift). In the social-context models, sphericity was violated for all effects involving context, except for the Context $\times$ Adaptation effect in Model D (face DTP). Therefore, Greenhouse-Geisser corrected

values were reported when $\epsilon < .75$ and Huynh-Feldt corrected values were reported when $\epsilon \geq .75$ (see full results in Appendix C, Tables C1-C4).

Single-person manipulations: Model A- loneliness rating. The statistical results showed that absolute loneliness rating shifts were significantly influenced by manipulation type (facial expression, weather, object, and background; $F(2.177, 115.395) = 146.544$, $p < .001$, $\eta^2p = .734$). Follow up paired samples t-tests comparing the four manipulation types showed that facial expression produced significantly larger absolute rating shifts than weather, object, and background (all Bonferroni-corrected $p < .001$). Weather also produced significantly larger absolute rating shifts than object and background (both Bonferroni-corrected $p < .001$). However, the comparison between object and background was not significant after Bonferroni correction.

Signed rating shifts for the ten individual levels were additionally tested against zero. All four facial expression levels showed significant signed shifts from zero (all $p_{\text{bonf}} < .001$, all $|dz| > 1.6$), while neither object level did (both $p_{\text{bonf}} = 1$). Within the background and weather manipulations, only one level of each differed from baseline: background-sparse produced a significant positive shift, indicating higher loneliness ratings than baseline ($t(53) = 7.236$, $p_{\text{bonf}} < .001$, $dz = .985$) while background-cluttered did not ($t(53) = -1.587$, $p_{\text{bonf}} = 1$, $dz = -.216$), and weather depressing produced a significant positive shift ($t(53) = 9.746$, $p_{\text{bonf}} < .001$, $dz = 1.326$) while weather holiday did not ($t(53) = -1.748$, $p_{\text{bonf}} = .862$, $dz = -.238$).

Single-person manipulations: Model B- face DTP. Manipulation type also produced a significant effect on face DTP ($F(3, 159) = 43.189$, $p < .001$, $\eta^2p = .449$). Follow up paired samples t-tests revealed that facial expressions produced significantly higher face DTP than weather, object, and background (all Bonferroni-corrected $p < .001$). Object produced significantly lower face DTP than weather ($t(53) = -3.30$, $p_{\text{bonf}} = .0103$, $dz = -.45$) and background ($t(53) = -4.57$, $p_{\text{bonf}} < .001$, $dz = -.62$), while weather and background did not significantly differ from each other.

Condition level comparisons against baseline were also conducted for the ten individual levels. None of the facial expression levels differed significantly from baseline (all $p_{\text{bonf}} = 1$), while all object and weather levels showed significantly lower face DTP (all $p_{\text{bonf}} < .001$). However, the two background levels differed: background-cluttered fell significantly below baseline ($t(53) = -7.98$, $p_{\text{bonf}} < .001$, $dz = -1.09$), while background-sparse did not ($t(53) = -2.37$, $p_{\text{bonf}} = .22$, $dz = -0.32$).

Social-context manipulations: Model C- loneliness rating. Results showed a significant main effect of adaptation ($F(1, 53) = 298.096$, $p < .001$, $\eta^2p = .849$). Context also showed a significant effect, with a smaller effect size ($F(1.411, 74.786) = 4.393$, $p = .027$, $\eta^2p = .077$). However, the Context $\times$ Adaptation interaction was not significant ($F(1.742, 92.316) = 1.881$, $p = .163$, $\eta^2p = .034$). Post-hoc comparisons for the context factor showed that exclusion produced a significantly larger absolute loneliness rating shift than friends (mean difference = 4.27, $SE = 1.57$, $p_{\text{tk}} = .02$) but did not significantly differ from dyad (mean difference = 4.01, $SE = 2.03$, $p_{\text{tk}} = .13$). Dyad and friends also did not significantly differ from each other.

Signed shift tests showed that all three adapted conditions differed significantly from zero, with large absolute effect sizes, with opposite signs across context types (all $p_{\text{bonf}} < .001$, all $|dz| > 2$). Adapted dyad and friends produced significant negative shifts, indicating lower loneliness ratings than baseline, whereas adapted exclusion produced a significant positive shift, indicating higher loneliness ratings than baseline ($t(53) = 19.291$, $p_{\text{bonf}} < .001$, $dz = 2.625$). Among the non-adapted conditions, exclusion and friends shifts from baseline reached significance, while dyad-present did not ($t(53) = -1.464$, $p_{\text{bonf}} = .895$, $dz = -0.199$).

Social-context manipulations: Model D- face DTP. Results showed a significant main effect of adaptation ($F(1, 53) = 33.897$, $p < .001$, $\eta^2p = .39$), followed by context ($F(1.701, 90.150) = 7.604$, $p = .002$, $\eta^2p = .125$). The Context $\times$ Adaptation interaction did not reach significance ($F(2, 106) = 2.977$, $p = .055$, $\eta^2p = .053$). Post-hoc comparisons for the context factor showed that exclusion produced significantly higher face DTP than friends (mean difference = 3.67, $SE = 1.03$, $p_{\text{tk}} = .002$), while dyad did not significantly differ from either exclusion or friends. Condition level comparisons showed that all six social-context levels had significantly lower face DTP than baseline (all $p_{\text{bonf}} < .001$).

The full ANOVA results, including all F-values, p-values, and partial eta squared (η^2p) values are provided in Appendix A, Tables A2–A5. The full follow-up results, including the pairwise t-tests, baseline paired-sample t-tests, and post-hoc comparisons, are provided in Appendix A, Tables A6–A14.

4.3 Additional observations

For descriptive visualization of gaze distribution, heatmaps were generated. As can be seen in Figure 6 panels (1) and (2), when participants did not look at the foreground person’s face, they looked mostly at the objects situated in the scene. A similar behavior can be seen in panels (3) – (8) of Figure 6, where participants mainly shifted their gaze toward other people’s faces whenever they were not looking at the foreground person’s face. However, the heatmaps showed that the faces of the other people in the friends and exclusion conditions received lower gaze density than both the foreground person’s face and the other person’s face in the dyad condition.



Figure 6: Descriptive gaze heatmaps for selected object, dyad, friends, and exclusion conditions. The figure shows representative heatmaps for eight stimulus images: (1) object non-social, (2) object social, (3) dyad present, (4) dyad present adapted, (5) friends present, (6) friends present adapted, (7) others excluding, and (8) others excluding adapted. Heatmaps were generated for each stimulus image by gathering filtered gaze coordinates (as described in section 3.5) across participants and trials. The gaze coordinates were converted from screen to image coordinates and accumulated into a density map, which was divided by the total number of gaze samples for that image and then smoothed with a Gaussian kernel. The smoothed maps were normalized using a single maximum value calculated across all stimulus images, allowing the heatmap colors to follow a shared scale across images, and a gamma correction was applied to improve the visibility of lower density regions. The color-coded density maps were overlaid on the original stimulus image, with warmer colors (red) indicating higher gaze density and cooler colors (blue) indicating lower gaze density.

5. Discussion

This study investigated which visual scene elements influence observers' judgments of a depicted foreground person's loneliness, and whether the same elements also shape gaze toward that person's face. The central finding is that the experimental manipulations produced partially overlapping but nonparallel patterns in judgment and face directed gaze. Some cues were associated with both large mean absolute loneliness rating shifts and relatively high face DTP, whereas for other cues, the statistical evidence was clearer for one outcome than for the other. This pattern is visible descriptively in Figure 5, where the conditions are distributed across the attention-judgment space rather than falling along a common axis.

Facial expression was the manipulation type for which judgment and gaze aligned most clearly, producing both the largest mean absolute rating shift and the highest face DTP among the four single-person manipulation types. The condition level signed shift tests further showed that all four facial expression levels produced significant directional shifts from baseline: depressed and socially disconnected expressions increased loneliness ratings, whereas happy and socially connected expressions decreased them. Although none of the facial expression levels differed significantly from baseline in face DTP, the baseline condition itself already showed high face DTP, consistent with its design as a neutral scene containing a single foreground person with a clearly visible neutral face. Facial expression levels therefore did not significantly change, rather than increase, this already high level of face directed attention. This is consistent with research showing that observers rely strongly on facial information when inferring another person's internal state (Wieser & Brosch, 2012; Jaeger et al., 2022), and with the face prioritization literature identifying faces as an important target of attention in social scenes (Birmingham et al., 2008; Dalmaso et al., 2025; Jaeger et al., 2022). Together, these findings suggest that manipulating facial expressions produced clear directional changes in loneliness ratings while maintaining the already high level of face directed attention observed at baseline.

The object manipulation showed the opposite profile. It produced the lowest face DTP of the four manipulation types, while its mean absolute rating shift did not differ significantly from background, the intended near control condition. At the condition level, comparisons with baseline showed that neither object level produced a significant signed shift in loneliness ratings, whereas both showed significantly lower face DTP. The object manipulation therefore showed clearer evidence of reducing face directed gaze than of producing a consistent directional change in perceived loneliness. One possible explanation is that the objects' visual salience or semantic inconsistency drew attention away from the foreground person's face and toward the objects themselves, as suggested by the heatmaps (seen in Figure 6), without providing sufficiently consistent social meaning to shift loneliness ratings in one direction (Schubö, 2009; Martinez-Cedillo & Foulsham, 2024).

The weather manipulation produced the second largest mean absolute rating shift among the four single-person manipulation types. However, the condition level signed shift tests showed that depressing weather significantly increased loneliness ratings, whereas holiday weather did not produce a significant directional shift (descriptively visualized in Figure 3). This suggests that the manipulation level effect was mainly driven by the depressing weather level rather than both levels. These results are partly consistent with research suggesting that atmospheric cues can influence emotional and social perception (Mei et al., 2023). For face DTP, both weather levels showed significant reductions relative to baseline, contrary to H2b. The weather manipulation was designed as a diffuse manipulation that did not introduce a discrete element expected to compete with the face and was therefore not predicted to significantly reduce face directed gaze. One possible explanation lies in a feature of the design: facial expression is the only manipulation that changed the same region used to compute face DTP. For all other manipulations the manipulated information lies outside the face AOI, so participants instructed to consider all available information when making their judgment may have sampled areas beyond the face to interpret the scene. Therefore, the reduction in face DTP for weather could reflect information sampling despite the absence of a discrete competing element (seen in Figure 6). Weather therefore showed an asymmetric pattern across judgment and gaze: both levels reduced face directed attention, whereas only depressing weather reliably shifted loneliness ratings.

The background levels also showed an asymmetric pattern across the two outcomes (descriptively visualized in Figures 3 and 4). Background sparse significantly increased loneliness ratings relative to baseline but did

not differ significantly from it in face DTP. In contrast, background cluttered did not produce a significant directional shift in loneliness ratings but showed significantly lower face DTP than baseline. Thus, background sparse showed clearer evidence of a change in judgment, whereas background cluttered showed clearer evidence of a change in face directed gaze. A possible explanation is that a cluttered background contains many small objects, so it may have competed for gaze in a similar way as the object manipulation did (Pedziwiatr et al., 2022; Schubö, 2009), whereas a sparse environment may suggest isolation because it contains fewer signs of social interaction, shifting judgment without providing a discrete element to attract gaze. These are two possible rather than tested explanations, but they suggest that the two background levels may have operated through different mechanisms rather than as equal sides of one manipulation, and that neither behaved as the near control condition the design intended.

In the social-context scenes, adaptation had a much larger effect size than context type on mean absolute loneliness rating shift. Thus, the magnitude of rating change was more strongly associated with whether the foreground person responded to the social situation than with context type. This is broadly consistent with research on expression-context congruence, where a match between a person's expression, bodily or facial, and the surrounding social context shaped both how the expression was recognized (Kret & De Gelder, 2010) and how the individual was judged (Chanes et al., 2018). The signed shift tests added directional detail: adapted dyad and friends decreased loneliness ratings relative to baseline, whereas adapted exclusion increased them. One possible explanation for this pattern is that positive engagement in the dyad and friends scenes could signal social inclusion, whereas the negative response to exclusion could reinforce the impression of social disconnection. The increase in loneliness ratings for adapted exclusion may therefore reflect the established conceptual link between exclusion and loneliness, both of which involve a threat to the need to belong (Basset et al., 2025; DeWall et al., 2009; Spithoven et al., 2017). Adaptation was also associated with lower face DTP when averaged across contexts. One possible explanation is that adaptation resolves the scene's social meaning. An adapted person makes the social situation less ambiguous, requiring less sustained inspection of the face. In contrast, a neutral person creates tension between expression and context, prompting repeated inspection of the face to gather enough information to resolve the scene's meaning. In addition, changes in the foreground person's gaze, head orientation, and posture in the adapted scenes may have directed attention toward the surrounding people and the wider interaction, which possibly explains the lower face DTP. Overall, adaptation was associated with larger absolute changes in loneliness ratings and lower face DTP.

On the other hand, context type showed a more limited pattern than adaptation. Averaged across adaptation levels, exclusion produced a larger mean absolute loneliness rating shift and higher face DTP than friends. However, neither outcome clearly separated dyad from the other two contexts. The findings therefore did not establish a graded ordering of exclusion, dyad, and friends in either loneliness judgment or face directed attention. Moreover, neither Context $\times$ Adaptation interaction reached significance, providing no clear evidence that the size of the adaptation effect varied across context types for either outcome. Overall, context type produced a partial distinction between exclusion and friends but did not clearly differentiate all three social contexts.

Taken together, the findings revealed three judgment–gaze patterns (summarized in Table A14). First, some manipulations shifted loneliness judgments without reducing face DTP: depressed and socially disconnected expressions and a sparse background increased loneliness, whereas happy and socially connected expressions decreased it. Second, other manipulations shifted judgment while also reducing face DTP: depressing weather, both exclusion conditions, and non-adapted friends increased loneliness, whereas adapted friends and adapted dyad decreased it. Finally, several manipulations reduced face DTP without producing a significant directional shift in loneliness ratings, including holiday weather, both object levels, a cluttered background, and non-adapted dyad. Overall, the findings suggest that cues that attracted or maintained attention to the face were not always the cues that most strongly influenced loneliness judgments. This suggests that perceived loneliness reflects facial, environmental, and social information rather than attention to the face alone. However, the way observers combined these cues was not directly examined here.

6. Limitations and future work

Several limitations of the present study should be acknowledged, including those related to the analytical approach, stimulus design, and questions that were not addressed in the current study.

One important limitation concerns the analysis. The use of absolute rating shift in the primary judgment analyses allows capturing the magnitude of change in perceived loneliness but not its direction. This choice enabled conducting inferential analysis through a single model across manipulations that pushed ratings in opposite directions. However, the trade-off is that this choice reduced the ability to distinguish whether a given cue made the foreground person appear less or more lonely. It also obscured directional variability between participants, meaning that a large mean absolute shift could reflect either a consistent change in one direction or highly variable changes in opposite directions. Condition level signed shift tests were used to determine the direction of rating change relative to baseline, while Figure 3 displays these directional patterns descriptively. A second limitation in the inferential analysis is that each manipulation type was represented by an average of different numbers of levels. Three of the manipulation types contained two levels each; while facial expressions contained four. Since we averaged across levels, the means being compared were built from different amounts of data, and a single level could influence a two-level average more than a four-level average. The comparison therefore may not fully separate the effect of the manipulation type itself from the effect of the levels representing it. The follow-up comparisons at the individual condition level partly address this, as they show how each level behaved relative to baseline.

A further limitation is that throughout the analyses we have averaged rating shift and face DTP over the six scenes. This was useful for testing the general effect of each manipulation type, but it also means that scene-level variability was not modeled directly. Therefore, some effects may partly reflect specific features of individual scene-condition images rather than the intended manipulation alone.

Beyond these measurement choices, several limitations also concern stimulus design. Generating stimuli using AI enabled systematic element level control that motivated the design. However, while the images were designed to be photorealistic, the outputs are synthetic and participants may perceive them differently from real photographs. Subtle artifacts or unrealistic visual details could influence both loneliness judgments and gaze behavior. At the end of the study, participants were asked in the Qualtrics questionnaire whether they noticed that the images were AI-generated. However, these responses were not formally analyzed in the present study. Analyzing them would be a useful next step, as it could help determine whether awareness of the synthetic nature of the images influenced participants' ratings or viewing patterns. Beyond the synthetic nature of the stimuli, our images were designed to be visually complex but remain static and are screen-based representations rather than real environments that participants could act within or move through. Prior work has found that gaze and attention can differ between viewing a person or scene on a screen and encountering it in real life, particularly for social stimuli (Pasqualetto & Kulke, 2023). Therefore, the gaze patterns observed here may not fully generalize to how attention is deployed during real-world viewing.

A further limitation of the stimulus set is that each scene set contained 17 images of the same foreground person with broadly similar scene compositions. Repeated exposure may have increased familiarity with the faces and stimulus structure, made the study's purpose more apparent, or encouraged participants to judge each image relative to previously viewed conditions. Ratings may also have changed over the course of the experiment as participants may have developed or recalibrated their internal scale for judging loneliness as they viewed more images. Because familiarity and trial-order effects were not examined, these possibilities remain speculative.

The results also suggest that some condition levels may not have been as comparable as the design intended. This was especially visible in the background manipulation, where sparse and cluttered backgrounds did not seem to function as two balanced levels of the same effect. Instead, the background levels produced asymmetrical patterns in both rating shift and face DTP (see Table A14), suggesting that they may have represented more than opposite ends of background density, for example by differing in visual salience, semantic content, or the social meaning implied by the environment. The adaptation manipulation raises a related problem; its two levels differed in more than the intended dimension. In non-adapted images the foreground person was always neutral, whereas in adapted images the foreground person appeared

positively engaged in the dyad and friends scenes but negatively affected in the exclusion scenes (seen in Figure 1). Adaptation, therefore, changed expression valence and intensity, as well as head and gaze direction and possibly posture, and did so in opposite directions across contexts. This means that the significant adaptation effect on both outcomes may partly reflect the combined effects of changes in expression valence and intensity, gaze direction, head orientation, and posture, rather than person–context congruence alone.

Another design limitation could be identified in the diffused manipulations. Weather and background were designed not to introduce discrete competing visual elements and were therefore not expected to strongly change gaze behavior. However, the significant reduction in face DTP compared to baseline (see Table A12) suggests either that the design did not fully isolate these manipulations as diffuse effects, or that diffuse manipulations can alter gaze behavior more than expected. The task instructions may have influenced this gaze pattern as participants were explicitly asked to consider several scene elements when judging loneliness. This may have encouraged deliberate scanning of the wider scene rather than spontaneous viewing. Therefore, lower face DTP may partly reflect task directed information sampling rather than stimulus driven attention alone. An additional unresolved finding concerns the dyad context. Dyad was expected to produce relatively high face DTP because it included only one additional person and therefore less attentional competition than friends or exclusion. However, the dyad conditions did not show the highest face DTP as predicted and did not differ significantly from either other context (see Table A9). One possible explanation is suggested by the gaze heatmaps (see Figure 6). In the dyad scenes, the other person’s face attracted about as much gaze as the foreground person’s face, whereas in the friends and exclusion scenes other people’s faces received less gaze density. Thus, despite containing fewer people, the dyad scenes included a strong competing face that may have divided attention between the two individuals.

Finally, the present dataset also leaves room for additional analyses. Because the eye-tracking analysis defined an area of interest only for the foreground person’s face, conclusions can be made about changes in face DTP, but not about where observers looked when they were not looking at the face. Interpretations that gaze was captured by other manipulated cues, such as objects, or by other people present, can therefore only be inferred descriptively from the heatmaps in Figure 6 rather than tested statistically. Considering that the raw gaze data already contains this information, a useful next step is to re-analyze this dataset with AOIs defined for all manipulated regions and for co-present people, to better establish which visual cues observers attend to when judging the loneliness of the foreground person.

7. Conclusion

To summarize the hypothesis outcomes, mean absolute loneliness rating shift followed the predicted ordering, supporting H1. H2a was also supported, as the predicted face DTP ordering was observed, but contrary to H2b prediction both weather levels and cluttered background also showed lower face DTP relative to baseline. H3 was partially supported: adapted conditions produced larger shifts than non-adapted ones, but exclusion exceeded only friends, not dyad. H4 was not supported, as dyad did not produce higher face DTP than the other social-context types.

Taken together, although this study did not directly test whether gaze was associated with loneliness judgments, comparing the findings across conditions revealed recurring patterns in how the two outcomes varied across the visual manipulations. Within the AI-generated static scenes used in this study, visual manipulations influenced both perceived loneliness and gaze toward the foreground person’s face, but the two outcomes did not follow a one-to-one relationship. Changes in facial expression produced the clearest directional changes in loneliness ratings while maintaining attention to the face. Atmospheric manipulations more often reduced face DTP, yet only some produced significant directional shifts in loneliness judgments. Background-density manipulations showed asymmetric judgment–gaze effects across their two levels, whereas the object manipulations reduced face DTP without producing significant directional shifts in loneliness judgments. In the social-context scenes, exclusion increased loneliness ratings in both adapted and non-adapted versions, whereas positive engagement in adapted dyad and friends was associated with lower ratings. Overall, face DTP indicated how much viewing time was directed toward the face, but it did not consistently indicate whether the depicted person would be judged as more or less lonely.

References

- Akram, U., & Stevenson, J. C. (2023). Altered perception of emotional faces in young adults experiencing loneliness after controlling for symptoms of insomnia, anxiety and depression. *Journal of Affective Disorders Reports*, 12, Article 100581. <https://doi.org/10.1016/j.jadr.2023.100581>
- Basset, G., Testa, A., Turati, C., Quadrelli, E., & Bulf, H. (2025). Ostracism affects children's behavioral reactivity and gaze cueing of attention. *PLOS ONE*, 20(3), Article e0320338. <https://doi.org/10.1371/journal.pone.0320338>
- Bell, J. L. (2014, April 4). *Human presence*. Jessica Louise Bell Photography. <https://jessicalouisebell.wordpress.com/2014/04/04/human-presence/>
- Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Social attention and real-world scenes: The roles of action, competition and social content. *The Quarterly Journal of Experimental Psychology*, 61(7), 986–998. <https://doi.org/10.1080/17470210701410375>
- Burgin, C. J., Brown, L. H., Royal, A., Silvia, P. J., Barrantes-Vidal, N., & Kwapil, T. R. (2012). Being with others and feeling happy: Emotional expressivity in everyday life. *Personality and Individual Differences*, 53(3), 185–190. <https://doi.org/10.1016/j.paid.2012.03.006>
- Cacioppo, J. T., & Hawkey, L. C. (2009). Perceived social isolation and cognition. *Trends in Cognitive Sciences*, 13(10), 447–454. <https://doi.org/10.1016/j.tics.2009.06.005>
- Caffier, P. P., Erdmann, U., & Ullsperger, P. (2003). Experimental evaluation of eye-blink parameters as a drowsiness measure. *European Journal of Applied Physiology*, 89, 319–325. <https://doi.org/10.1007/s00421-003-0807-5>
- Chanes, L., Wormwood, J. B., Betz, N., & Barrett, L. F. (2018). Facial expression predictions as drivers of social perception. *Journal of Personality and Social Psychology*, 114(3), 380–396. <https://doi.org/10.1037/pspa0000108>
- Dalmaso, M., Gobbini, M. I., Ricciardelli, P., Ritchie, K. L., & Pecchinenda, A. (2025). Face perception: A window into the social mind [Editorial]. *Scientific Reports*, 15, Article 32362. <https://doi.org/10.1038/s41598-025-17861-x>
- Danneel, S., Nelemans, S., Spithoven, A., Bastin, M., Bijttebier, P., Colpin, H., van Den Noortgate, W., van Leeuwen, K., Verschueren, K., & Goossens, L. (2019). Internalizing problems in adolescence: Linking loneliness, social anxiety symptoms, and depressive symptoms over time. *Journal of Abnormal Child Psychology*, 47(10), 1691–1705. <https://doi.org/10.1007/s10802-019-00539-0>
- DeWall, C. N., Maner, J. K., & Rouby, D. A. (2009). Social exclusion and early-stage interpersonal perception: Selective attention to signs of acceptance. *Journal of Personality and Social Psychology*, 96(4), 729–741. <https://doi.org/10.1037/a0014634>
- De Winter, J. C. F. (2026). ai-context-valence-loneliness. GitHub Repository. <https://github.com/JoostdeWinter/ai-context-valence-loneliness>
- Hawkey, L. C., & Cacioppo, J. T. (2010). Loneliness matters: A theoretical and empirical review of consequences and mechanisms. *Annals of Behavioral Medicine*, 40(2), 218–227. <https://doi.org/10.1007/s12160-010-9210-8>

- Henderson, J. M., Chanceaux, M., & Smith, T. J. (2009). The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements. *Journal of Vision*, 9(1), Article 32, 1–8. <https://doi.org/10.1167/9.1.32>
- Henderson, J. M. (2017). Gaze control as prediction. *Trends in Cognitive Sciences*, 21(1), 15–23. <https://doi.org/10.1016/j.tics.2016.11.003>
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Hughes, M. E., Waite, L. J., Hawkey, L. C., & Cacioppo, J. T. (2004). A short scale for measuring loneliness in large surveys: Results from two population-based studies. *Research on Aging*, 26(6), 655–672. <https://doi.org/10.1177/0164027504268574>
- Jaeger, B., Evans, A. M., Stel, M., & van Beest, I. (2022). Understanding the role of faces in person perception: Increased reliance on facial appearance when judging sociability. *Journal of Experimental Social Psychology*, 100, Article 104288. <https://doi.org/10.1016/j.jesp.2022.104288>
- Ju, Q., Xu, Z., Chen, Z., Fan, J., Zhang, H., & Peng, Y. (2025). Screening social anxiety with the Social Artificial Intelligence Picture System. *Journal of Anxiety Disorders*, 109, Article 102955. <https://doi.org/10.1016/j.janxdis.2024.102955>
- Klimstra, T. A., Frijns, T., Keijsers, L., Denissen, J. J. A., Raaijmakers, Q. A. W., van Aken, M. A. G., Koot, H. M., van Lier, P. A. C., & Meeus, W. H. J. (2011). Come rain or come shine: Individual differences in how weather affects mood. *Emotion*, 11(6), 1495–1499. <https://doi.org/10.1037/a0024649>
- Kret, M. E., & De Gelder, B. (2010). Social context influences recognition of bodily expressions. *Experimental Brain Research*, 203(1), 169–180. <https://doi.org/10.1007/s00221-010-2220-8>
- Krumhuber, E. G., Skora, L. I., Hill, H. C. H., & Lander, K. (2023). The role of facial movements in emotion recognition. *Nature Reviews Psychology*, 2(5), 283–296. <https://doi.org/10.1038/s44159-023-00172-1>
- Mack, A., & Rock, I. (1998). *Inattention blindness*. MIT Press.
- Martinez-Cedillo, A. P., & Foulsham, T. (2024). Don't look now! Social elements are harder to avoid during scene viewing. *Vision Research*, 216, 108356. <https://doi.org/10.1016/j.visres.2023.108356>
- McDowell, C. P., Meyer, J. D., Russell, D. W., Brower, C. S., Lansing, J., & Herring, M. P. (2021). Bidirectional associations between depressive and anxiety symptoms and loneliness during the COVID-19 pandemic: Dynamic panel models with fixed effects. *Frontiers in Psychiatry*, 12, Article 738892. <https://doi.org/10.3389/fpsy.2021.738892>
- Mei, Y., Xu, L., & Li, Z. (2023). Study on emotional perception of Hangzhou West Lake Scenic Area in spring under the influence of meteorological environment. *International Journal of Environmental Research and Public Health*, 20(3), Article 1905. <https://doi.org/10.3390/ijerph20031905>
- Muhammad, L. N. (2023). Guidelines for repeated measures statistical analysis approaches with basic science research considerations. *The Journal of Clinical Investigation*, 133(11), Article e171058. <https://doi.org/10.1172/JCI171058>
- Mushtaq, R., Shoib, S., Shah, T., & Mushtaq, S. (2014). Relationship between loneliness, psychiatric disorders and physical health? A review on the psychological aspects of loneliness. *Journal of Clinical and Diagnostic Research*, 8(9), WE01–WE04. <https://doi.org/10.7860/JCDR/2014/10077.4828>

- Papadopoulos, N., Navaneethan, S., Bai, S., Samanta, A., & Sajda, P. (2026). Gaze patterns predict preference and confidence in pairwise AI image evaluation. In *Proceedings of the 2026 Symposium on Eye Tracking Research and Applications (ETRA '26)*. Association for Computing Machinery. <https://doi.org/10.1145/3797246.3803031>
- Pärnamets, P., Johansson, R., Gidlöf, K., & Wallin, A. (2016). How information availability interacts with visual attention during judgment and decision tasks. *Journal of Behavioral Decision Making*, 29(2–3), 218–231. <https://doi.org/10.1002/bdm.1902>
- Pasqualetto, L., & Kulke, L. (2023). Effects of emotional content on social inhibition of gaze in live social and non-social situations. *Scientific Reports*, 13, Article 14151. <https://doi.org/10.1038/s41598-023-41154-w>
- Pedziwiatr, M. A., Kümmerer, M., Wallis, T. S. A., Bethge, M., & Teufel, C. (2022). Semantic object-scene inconsistencies affect eye movements, but not in the way predicted by contextualized meaning maps. *Journal of Vision*, 22(2), Article 9. <https://doi.org/10.1167/jov.22.2.9>
- Perlman, D., & Peplau, L. A. (1981). Toward a social psychology of loneliness. In S. Duck & R. Gilmour (Eds.), *Personal relationships in disorder* (pp. 31–56). Academic Press.
- Qualter, P., Vanhalst, J., Harris, R., van Roekel, E., Lodder, G., Bangee, M., Maes, M., & Verhagen, M. (2015). Loneliness across the life span. *Perspectives on Psychological Science*, 10(2), 250–264. <https://doi.org/10.1177/1745691615568999>
- Renier, L. A., Shubham, K., Vijay, R. S., Mishra, S. S., Kleinlogel, E. P., Jayagopi, D. B., & Schmid Mast, M. (2024). A deepfake-based study on facial expressiveness and social outcomes. *Scientific Reports*, 14, Article 3642. <https://doi.org/10.1038/s41598-024-53475-5>
- Rokach, A., & Berman, D. (2025). Loneliness: An existential and public health issue [Editorial]. *International Journal of Environmental Research and Public Health*, 22(10), Article 1544. <https://doi.org/10.3390/ijerph22101544>
- Schubö, A. (2009). Salience detection and attentional capture. *Psychological Research*, 73(2), 233–243. <https://doi.org/10.1007/s00426-008-0215-x>
- Skaramagkas, V., Giannakakis, G., Ktistakis, E., Manousos, D., Karatzanis, I., Tachos, N. S., Tripoliti, E., Marias, K., Fotiadis, D. I., & Tsiknakis, M. (2023). Review of eye tracking metrics involved in emotional and cognitive processes. *IEEE Reviews in Biomedical Engineering*, 16, 260–277. <https://doi.org/10.1109/RBME.2021.3066072>
- Spithoven, A. W. M., Bijttebier, P., & Goossens, L. (2017). It is all in their mind: A review on information processing bias in lonely individuals. *Clinical Psychology Review*, 58, 97–114. <https://doi.org/10.1016/j.cpr.2017.10.003>
- Ting, C.-C., & Gluth, S. (2024). Unraveling information processes of decision-making with eye tracking data. *Frontiers in Behavioral Economics*, 3, Article 1384713. <https://doi.org/10.3389/frmbe.2024.1384713>
- van Berlo, Z. M. C., Campbell, C., & Voorveld, H. A. M. (2024). The MADE framework: Best practices for creating effective experimental stimuli using generative AI. *Journal of Advertising*, 53(5), 732–753. <https://doi.org/10.1080/00913367.2024.2397777>
- Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering psychology and human performance* (4th ed.). Pearson.

Wieser, M. J., & Brosch, T. (2012). Faces in context: A review and systematization of contextual influences on affective face processing. *Frontiers in Psychology*, 3, Article 471.
<https://doi.org/10.3389/fpsyg.2012.00471>

Appendix A

Table A1. Descriptive statistics for face dwell time percentage (DTP) (%)

Condition levels	Mean	SD	95% CI
baseline	63.01	13.39	[59.35, 66.66]
Face depressed	63.18	15.19	[59.03, 67.32]
Face happy	61.55	14.51	[57.59, 65.51]
Face socially connected	65.05	13.53	[61.36, 68.75]
Face socially disconnected	66.34	16.17	[61.93, 70.75]
Weather depressing	53.42	13.80	[49.65, 57.19]
Weather holiday	56.44	13.06	[52.88, 60.01]
Object nonsocial	49.56	11.18	[46.51, 52.61]
Object social	52.20	12.35	[48.83, 55.57]
Background cluttered	52.29	13.38	[48.64, 55.94]
Background sparse	59.76	13.11	[56.18, 63.34]
Dyad present	43.04	9.65	[40.41, 45.68]
Dyad present adapted	40.45	9.33	[37.90, 43.00]
Friends present	42.04	10.08	[39.29, 44.80]
Friends present adapted	38.18	9.56	[35.57, 40.78]
Others excluding	47.23	10.45	[44.38, 50.09]
Others excluding adapted	40.32	11.90	[37.08, 43.57]

Table A2. Model A: repeated-measures ANOVA results for: | rating shift| ~ Single-person Manipulations

Condition	df1	df2	F (df1, df2)	p-value	η^2p	ϵ
Single-person Manipulations (facial expression, object, weather, background)	2.177	115.395	146.544	<0.001	0.734	0.726

Note: Greenhouse–Geisser corrected p-values were used.

Table A3. Model B: repeated-measures ANOVA results for: face DTP ~ Single-person Manipulations

Condition	df1	df2	F (df1, df2)	p-value	η^2p	ϵ
Single-person Manipulations (facial expression, object, weather, background)	3	159	43.189	<0.001	0.449	-

Note: Model met sphericity and was left uncorrected.

Table A4. Model C: repeated-measures ANOVA results for: | rating shift| ~ Context × Adaptation interaction.

Condition	df1	df2	F (df1, df2)	p-value	η^2p	ϵ
Context	1.411	74.786	4.393	0.027	0.077	0.706
Adaptation	1	53	298.096	<0.001	0.849	-
Context × adaptation	1.742	92.316	1.881	0.163	0.034	0.871

Note: Greenhouse–Geisser corrected p-values were used for effects involving the three-level context factor, and Huynh-Feldt corrected p-values for the context × adaptation interaction.

Table A5. Model D: repeated-measures ANOVA results for: face DTP ~ Context × Adaptation interaction.

Condition	df1	df2	F (df1, df2)	p-value	η^2p	ϵ
Context	1.701	90.150	7.604	0.002	0.125	0.850
Adaptation	1	53	33.897	<0.001	0.39	-
Context × adaptation	2	106	2.977	0.055	0.053	-

Note: Huynh-Feldt corrected p-values were used for effects involving the context main effect; the interaction met sphericity and was left uncorrected.

Table A6. Single-person manipulations pairwise t-tests for absolute rating shift

Condition 1	Condition 2	t	Degrees of freedom	p-value	Bonferroni-corrected p-values	Cohen's dz
face	weather	10.79	53	<0.001	<0.001	1.47
face	object	16.49	53	<0.001	<0.001	2.24
face	background	15.09	53	<0.001	<0.001	2.05

object	background	-2.09	53	0.04	0.25	-0.29
weather	object	5.69	53	<0.001	<0.001	0.77
weather	background	4.77	53	<0.001	<0.001	0.65

Table A7. Single-person manipulations pairwise t-tests for face DTP

Condition 1	Condition 2	t	Degrees of freedom	p-value	Bonferroni-corrected p-values	Cohen's dz
face	weather	7.14	53	<0.001	<0.001	0.97
face	object	9.64	53	<0.001	<0.001	1.31
face	background	7.21	53	<0.001	<0.001	0.98
object	weather	-3.30	53	0.0017	0.0103	-0.45
object	background	-4.57	53	<0.001	0.00018	-0.62
weather	background	-1.13	53	0.2637	1	-0.15

Table A8. Social-context manipulations post-hoc comparisons for absolute rating shift for the context levels only

Context 1	Context 2	Mean difference	SE	Tukey–Kramer-corrected p-values	95% CI
Exclusion	Dyad	4.01	2.03	0.13	[-0.89, 8.92]
Exclusion	Friends	4.27	1.57	0.02	[0.48, 8.06]
Dyad	Friends	0.26	1.11	0.97	[-2.41, 2.93]

Table A9. Social-context manipulations post-hoc comparisons for face DTP for the context levels only

Context 1	Context 2	Mean difference	SE	Tukey–Kramer-corrected p-values	95% CI
Exclusion	Dyad	2.03	1.06	0.14	[-0.52, 4.58]
Exclusion	Friends	3.67	1.03	0.002	[1.18, 6.16]
Dyad	Friends	1.64	0.69	0.06	[-0.04, 3.31]

Table A10. Single-person manipulations one-sample t-tests for signed rating shift vs. baseline, individual condition levels

Condition levels	Mean of signed shift	t	Degrees of freedom	Bonferroni-corrected p-values	Cohen's dz
Face depressed	27.114	14.505	53	<0.001	1.974
Face happy	-23.388	-14.285	53	<0.001	-1.944
Face socially connected	-19.265	-13.842	53	<0.001	-1.884
Face socially disconnected	18.777	11.780	53	<0.001	1.603
Weather depressing	12.689	9.746	53	<0.001	1.326
Weather holiday	-2.178	-1.748	53	0.862	-0.238
Object non-social	1.051	1.123	53	1	0.153
Object social	-1.265	-1.442	53	1	-0.196
Background cluttered	-1.336	-1.587	53	1	-0.216
Background sparse	6.393	7.236	53	<0.001	0.985

Table A11. Social-context manipulations one-sample t-tests for signed rating shift vs. baseline, individual condition levels

Condition levels	Mean of signed shift	t	Degrees of freedom	Bonferroni-corrected p-values	Cohen's dz
Dyad present	-2.471	-1.464	53	0.895	-0.199
Dyad present adapted	-30.072	-15.994	53	<0.001	-2.176
Friends present	7.599	4.759	53	<0.001	0.648
Friends present adapted	-27.664	-14.921	53	<0.001	-2.031
Others excluding	11.433	8.164	53	<0.001	1.111
Others excluding adapted	34.454	19.291	53	<0.001	2.625

Table A12. Single-person manipulations paired-sample t-tests for face DTP vs. baseline, individual condition levels

Condition levels	Mean of level	Mean of baseline	t	Degrees of freedom	Bonferroni-corrected p-values	Cohen's dz
Face depressed	63.176	63.008	0.106	53	1	0.014
Face happy	61.551	63.008	-0.933	53	1	-0.127
Face socially connected	65.054	63.008	1.464	53	1	0.199
Face socially disconnected	66.338	63.008	1.934	53	0.585	0.263
Weather depressing	53.421	63.008	-5.740	53	<0.001	-0.781
Weather holiday	56.444	63.008	-4.648	53	<0.001	-0.633
Object non-social	49.560	63.008	-8.666	53	<0.001	-1.179
Object social	52.203	63.008	-8.279	53	<0.001	-1.127
Background cluttered	52.289	63.008	-7.982	53	<0.001	-1.086
Background sparse	59.757	63.008	-2.369	53	0.215	-0.322

Table A13. Social-context manipulations paired-sample t-tests for face DTP vs. baseline, individual condition levels

Condition levels	Mean of level	Mean of baseline	t	Degrees of freedom	Bonferroni-corrected p-values	Cohen's dz
Dyad present	43.042	63.008	-14.133	53	<0.001	-1.923
Dyad present adapted	40.451	63.008	-15.585	53	<0.001	-2.121
Friends present	42.043	63.008	-14.734	53	<0.001	-2.005
Friends present adapted	38.176	63.008	-18.620	53	<0.001	-2.534
Others excluding	47.234	63.008	-11.037	53	<0.001	-1.502

Others excluding adapted	40.324	63.008	-14.540	53	<0.001	-1.979
--------------------------	--------	--------	---------	----	--------	--------

Table A14. Summary of condition level effects on perceived loneliness and face DTP relative to baseline

Group	Condition level	Effect on loneliness rating	Effect on face DTP
Single-person	Face depressed	Increased loneliness	No significant change
Single-person	Face happy	Decreased loneliness	No significant change
Single-person	Face socially connected	Decreased loneliness	No significant change
Single-person	Face socially disconnected	Increased loneliness	No significant change
Single-person	Weather depressing	Increased loneliness	Reduced face DTP
Single-person	Weather holiday	No clear loneliness effect	Reduced face DTP
Single-person	Object non-social	No clear loneliness effect	Reduced face DTP
Single-person	Object social	No clear loneliness effect	Reduced face DTP
Single-person	Background cluttered	No clear loneliness effect	Reduced face DTP
Single-person	Background sparse	Increased loneliness	No significant change
Social-context	Dyad present	No clear loneliness effect	Reduced face DTP
Social-context	Dyad present adapted	Decreased loneliness	Reduced face DTP
Social-context	Friends present	Increased loneliness	Reduced face DTP
Social-context	Friends present adapted	Decreased loneliness	Reduced face DTP
Social-context	Others excluding	Increased loneliness	Reduced face DTP
Social-context	Others excluding adapted	Increased loneliness	Reduced face DTP

Note: Using Bonferroni-corrected $p < .05$ as the criterion, summarizing Tables A10 – A13.

Appendix B

Examining the loneliness ratings across the 17 condition levels, some levels showed higher average loneliness ratings, such as the “depressed face” and “others excluding adapted” levels. Other levels showed lower average loneliness ratings, including the “happy face”, “dyad present adapted”, and “friends present adapted” levels. Some levels appeared closer to baseline, including the object levels and the “background cluttered” level (see Figure B.1).

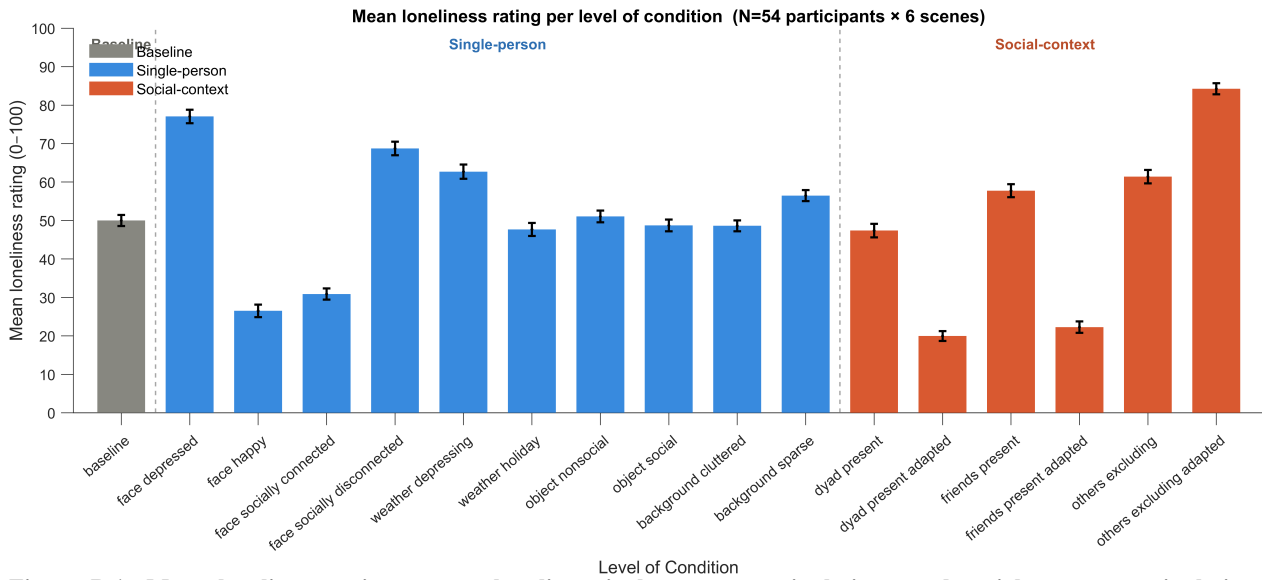


Figure B.1: Mean loneliness ratings across baseline, single-person manipulations, and social-context manipulations including all 17 condition levels. Bars show the mean loneliness rating on a 0–100 scale, where higher scores indicate greater perceived loneliness. Ratings were averaged across six scenes and 54 participants per condition level. The grey bar represents the baseline condition, blue bars represent single-person condition levels, and red bars represent social-context condition levels. Error bars indicate ± 1 Standard Error of the Mean (SEM). Vertical dashed lines separate the baseline, Single-person, and Social-context sections of the plot.

An additional descriptive scatter plot was used to visualize mean face DTP against mean loneliness absolute rating shift, relative to matched baseline, across manipulations (see Figure B.2). The plot summarizes the 10 overall conditions belonging to single-person and social-context manipulations (see Table 1 and Table 2). Among the single-person manipulations, the facial-expression condition showed relatively high face DTP together with a relatively large absolute rating shift. In contrast, adapted social-context conditions showed relatively low face DTP together with relatively large absolute rating shifts.

attention vs loneliness rating shift (10 condition cells) — 5151/5184 observations

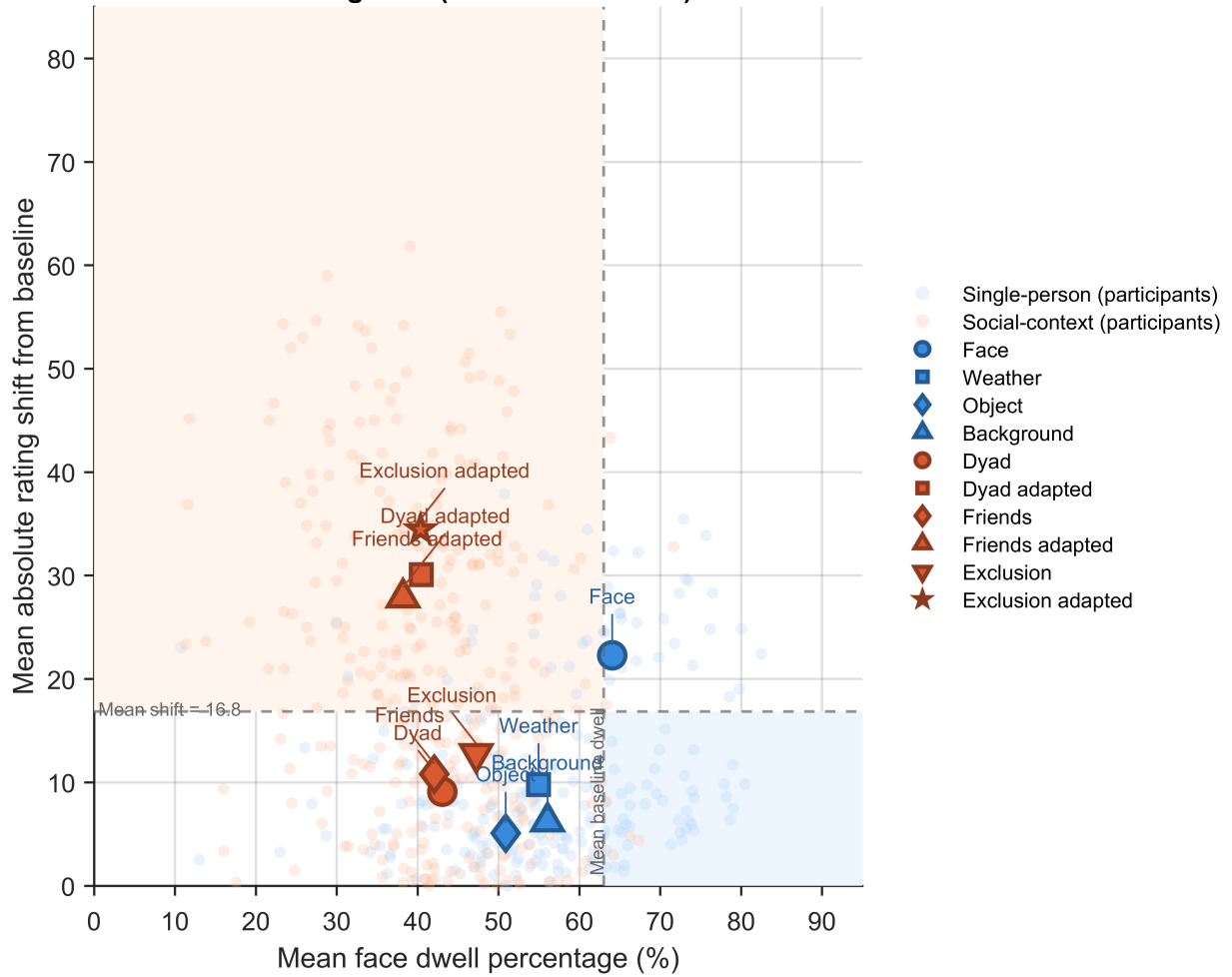


Figure B.2: Descriptive comparison between face dwell time percentage and absolute loneliness rating shift at the condition level. The scatter plot shows the relationship between mean face dwell time percentage (DTP) and mean absolute rating shift from baseline across the 10 experimental conditions. Large markers represent the group mean for each condition cell: four single-person scene manipulations in blue and six social-context manipulations in red. Faint background points show participant-level values: each point is one participant's mean face DTP and mean absolute rating shift for one condition (54 participants $\times$ 10 conditions). The x-axis represents mean face DTP, and the y-axis represents the absolute rating shift from the matched baseline scene, where higher values indicate a larger change in loneliness ratings regardless of direction. The vertical dashed line marks the mean face DTP across baseline scenes, and the horizontal dashed line marks the mean absolute rating shift across the 10 conditions. A total of 5151 out of 5184 expected observations (16 non-baseline condition levels $\times$ 6 scenes $\times$ 54 participants) were included after data filtering.

Appendix C

The assumptions of the repeated-measures ANOVAs were assessed before the statistical analysis. This included extreme outlier check, dependent variables normality check, and sphericity check for factors that had three or more levels, including manipulation type (facial expression, weather, object, background) for Models A and B, Context (dyad, friends, exclusion), and context $\times$ adaptation interaction for Models C and D. Adaptation was not checked for sphericity as it only has two levels.

Table C1: Normality checks for condition level values

Group	Model	Outcome	Condition	p-value	Decision
Single-person	Model A	Absolute rating shift	Facial expression	.127	No evidence of deviation from normality
Single-person	Model A	Absolute rating shift	Weather	.005	Deviation from normality
Single-person	Model A	Absolute rating shift	Object	.108	No evidence of deviation from normality
Single-person	Model A	Absolute rating shift	Background	.310	No evidence of deviation from normality
Single-person	Model B	Face DTP	Facial expression	.054	No evidence of deviation from normality
Single-person	Model B	Face DTP	Weather	> .500	No evidence of deviation from normality
Single-person	Model B	Face DTP	Object	> .500	No evidence of deviation from normality
Single-person	Model B	Face DTP	Background	.382	No evidence of deviation from normality
Social-context	Model C	Absolute rating shift	Dyad	.005	Deviation from normality
Social-context	Model C	Absolute rating shift	Dyad adapted	> .500	No evidence of deviation from normality
Social-context	Model C	Absolute rating shift	Friends	.017	Deviation from normality
Social-context	Model C	Absolute rating shift	Friends adapted	> .500	No evidence of deviation from normality
Social-context	Model C	Absolute rating shift	Exclusion	.150	No evidence of deviation from normality
Social-context	Model C	Absolute rating shift	Exclusion adapted	.406	No evidence of deviation from normality
Social-context	Model D	Face DTP	Dyad	.262	No evidence of deviation from normality
Social-context	Model D	Face DTP	Dyad adapted	.005	Deviation from normality
Social-context	Model D	Face DTP	Friends	.478	No evidence of deviation from normality
Social-context	Model D	Face DTP	Friends adapted	.283	No evidence of deviation from normality
Social-context	Model D	Face DTP	Exclusion	> .500	No evidence of deviation from normality
Social-context	Model D	Face DTP	Exclusion adapted	> .500	No evidence of deviation from normality

Table C2: Results of outlier checks using ± 3 SD criterion

Group	Model	Outcome	Condition	Number of possible outliers
Single-person	Model A	Absolute rating shift	Weather	1

Single-person	Model A	Absolute rating shift	Background	1
Single-person	Model B	Face DTP	Facial expression	1
Single-person	Model B	Face DTP	Background	1
Social-context	Model D	Face DTP	Dyad adapted	1

Table C3. Sphericity checks for Single-person manipulations repeated-measures ANOVAs

Model	Outcome	Mauchly's W	χ^2	Degree of freedom	p-value	Greenhouse–Geisser ϵ	Huynh–Feldt ϵ	Lower-bound ϵ	Decision
Model A	Absolute rating shift	0.506	35.228	5	< .001	0.726	0.758	0.333	Sphericity violated
Model B	Face DTP	0.853	8.225	5	.144	0.916	0.971	0.333	Sphericity not violated

Note: For absolute rating shift (Model A), Greenhouse–Geisser and Huynh–Feldt epsilon estimates straddled the conventional .75 threshold ($\epsilon = .726$ and $\epsilon = .758$, respectively); the Greenhouse–Geisser correction was applied as the more conservative choice in this borderline case.

Table C4: Sphericity checks for Social-context manipulations repeated-measures ANOVAs

Model	Outcome	Mauchly's W	χ^2	Degree of freedom	p-value	Greenhouse–Geisser ϵ	Huynh–Feldt ϵ	Lower-bound ϵ	Decision
Model C-Context	Absolute rating shift	0.583	28.091	2	< .001	0.706	0.719	0.5	Sphericity violated
Model C-Context $\times$ Adaptation	Absolute rating shift	0.818	10.448	2	0.0054	0.846	0.871	0.5	Sphericity violated
Model D-Context	Face DTP	0.789	12.324	2	.002	0.826	0.849	0.5	Sphericity violated
Model D-Context $\times$ Adaptation	Face DTP	0.983	0.904	2	0.636	0.983	1	0.5	Sphericity Not violated

Note: The Greenhouse–Geisser correction was used for absolute rating shift because $\epsilon < .75$, whereas the Huynh–Feldt correction was used for face DTP because $\epsilon > .75$. Effects involving only adaptation were not corrected because adaptation had two levels.