Exploring the Usability of the German COVID-19 Contact Tracing App in a Combined Eye Tracking and Retrospective Think Aloud Study

Michael Winter[1], Harald Baumeister[2], Ulrich Frick[3], Miles Tallon[3], Manfred Reichert[1], and Rüdiger Pryss[4]

Abstract— In the course of the corona virus (COVID-19) pandemic, many digital solutions for mobile devices (e.g., apps) were presented in order to provide additional resources supporting the control of the pandemic. Contact tracing apps (i.e., identify persons who may have been in contact with a COVID-19 infected) constitute one of the most popular as well as promising solutions. However, as a prerequisite for an effective application, such apps highly depend on being used by large numbers of the population. Consequently, it is important that these apps offer a high usability for everyone. We therefore conducted an exploratory study to learn more about the usability of the German COVID-19 contact tracing app Corona-Warn-App (CWA). More specifically, N = 15 participants assessed the CWA, relying on a combined eye tracking and retrospective think aloud approach. The results indicate, on the one hand, that the CWA leaves a promising impression for pandemic control, as essential functions are easily recognized. However, on the other hand, issues were revealed (e.g., privacy policy) that could be addressed in future updates more properly.

I. INTRODUCTION

The COVID-19 pandemic is currently impacting almost all areas of our personal lives around the globe. Our whole society is distinctly affected and challenged by the effects of the pandemic in an unprecedented manner. Quarantine and the adverse effects of isolation on the mental health [1], home schooling and emerging inequalities in education [2], abandoned culture activities [3], and tense situations in hospital intensive care units due to the increased number of COVID-19 patients [4] are some examples of the effects. For this reason, it is of utmost importance to provide collaborative, transparent, and scientifically substantiated approaches in order to control the COVID-19 pandemic.

In the pandemic context, the proliferation of mobile devices (e.g., smartphones, tablets) can play an essential role in its control [5]. Since a large part of the population uses such devices on a daily basis, they can make a substantial contribution especially in the rapid communication as well as dissemination of information during the pandemic. Used correctly and in large numbers, mobile devices can be depicted as potential game changers in this kind of situation (e.g., better overview of the COVID-19 situation) [6].

Consequently, many digital solutions have been released for mobile devices in the wake of the COVID-19 pandemic [7]. One of the popular apps in this context constitutes the German contact tracing app Corona-Warn-App (CWA) [8]. The latter informs users about a possible infection risk with the main objective of interrupting infection chains and, thus, preventing the spread of COVID-19. As a prerequisite for an effective application of the CWA, however, it must be used in large numbers by the population. However, despite the clear benefits of the CWA, the app is often the subject of critical debates and many (potential) users are still concerned about the usefulness (e.g., privacy policy) of the CWA.

Therefore, this paper presents the results of an exploratory study, in which the usability of the CWA was evaluated. More specifically, for N = 15 study participants, eye tracking was combined with a retrospective think aloud (RTA) approach to assess the usability of the CWA. The gained insights reveal well implemented essential functions as well as indicate opportunities for the improvement of the CWA in order to increase its usability regarding the pandemic. Notably, the study was conducted in the scope of the COMPASS project[*], which is part of the *Netzwerk Universitätsmedizin*.

The structure of this paper is as follows: After providing information about materials and methods of the study in Section II, results are analyzed and discussed in Section III, confronted study limitations are shown in the latter section as well. Finally, Section IV concludes the paper and gives an outlook on future work.

II. MATERIALS AND METHODS

A. Methods

We relied on eye tracking in this study as an established technique in usability research [9], [10]. Eye tracking describes the analysis of eye movements (e.g., fixation, saccades) in a static or dynamic stimulus (e.g., image, video). Further, eye tracking was combined with the retrospective think aloud (RTA) technique in the study. RTA constitutes a method to collect data about the reasoning and intents of an individual during a task and is widely adopted in usability testing [11], [12]. Prior research showed that the combination of eye tracking and RTA enables a fine-grained understanding of usability problems [13], [14]. In general, several approaches exist for think aloud (i.e., concurrent, retrospective,

[1]Michael Winter and Manfred Reichert are with the Institute of Databases and Information Systems, Ulm University, Ulm, Germany `{michael.winter,manfred.reichert}@uni-ulm.de`

[2]Harald Baumeister is with the Department of Clinical Psychology and Psychotherapy, Ulm University, Ulm, Germany `harald.baumeister@uni-ulm.de`

[3]Ulrich Frick and Miles Tallon are with the HSD Research Centre Cologne, HSD University of Applied Sciences, Cologne, Germany `{u.frick,m.tallon}@hs-doepfer.de`

[4]Rüdiger Pryss is with the Institute of Clinical Epidemiology and Biometry, University of Würzburg, Würzburg, Germany `ruediger.pryss@uni-wuerzburg.de`

ID	Gender	Age	Education	CWA	Other
1	1	28	1	0	–
2	0	32	1	0	–
3	1	34	1	1	–
4	1	31	2	0	–
5	0	25	1	1	–
6	0	29	1	1	CH, CC
7	1	31	1	0	–
8	0	28	1	1	CH
9	0	30	1	0	–
10	0	44	2	1	–
11	1	26	0	1	–
12	0	26	0	1	–
13	0	41	2	1	–
14	0	29	1	1	–
15	0	29	1	1	–

Note: Gender: 0 = male, 1 = female; Education: 0 = bachelor's degree, 1 = master's degree, 2 = doctor of philosophy; CWA: 0 = No, 1 = Yes; Other: CH = Corona Health, CC = Corona Check

and hybrid) [15]. However, research demonstrated that the concurrent think aloud (i.e., verbalization during a task) may lead to a more biased usability evaluation (e.g., more fixations than usual) [16]. For this reason, we decided to use a combination of eye tracking and RTA in this exploratory study. On the one hand, this allowed us to analyze the eye movements of the participants during the assessment of the CWA in order to identify distracting or conducive elements therein. On the other hand, the verbalization of the applied strategies from the participants in the RTA, which was applied after eye tracking, provided us with additional insights about assessment strategies.

B. Participants

The study enrolled a total of N = 15 participants, which were recruited at Ulm University (N = 8) and the University of Würzburg (N = 7). Five participants were female, 10 male and the mean (M) age was M = 30.87 years (standard deviation (SD) = 5.14). Regarding the highest educational qualification, 2 participants hold a bachelor's degree, 10 a master's degree, and 3 a doctor of philosophy. 10 participants stated that they had already installed the CWA on their own smartphone and 2 out of the 10 participants had additional COVID-19 tracing apps installed (i.e., Corona Health, Corona Check [17]). Table I summarizes the study participants (i.e., ID), thereby presenting gender, age, highest educational background, CWA installed, and other apps installed on the smartphone.

C. Materials

In this study, the German COVID-19 tracing App Corona-Warn-App (CWA) was evaluated with regard to the aspect of usability. The CWA is the most widespread contact tracing app in Germany, which even allows for contact tracing in many other countries of the European Union [8]. The CWA potentially constitutes a pivotal role in the control of the COVID-19 pandemic (e.g., track infection chains and interrupt the spread of COVID-19). For this reason, it is

of particular interest to investigate how users perceive the CWA's usability during the COVID-19 pandemic. In order to investigate the latter, the participants from the study were asked to find the following three functions (i.e., function-finding task (FFT)) in the CWA:

FFT 1: Where can I find my infection risk?

FFT 2: Where can I submit a positive COVID-19 test result?

FFT 3: Where can I find the privacy policy?

The first two functions represent the most important tools for the non-pharmaceutical control of the COVID-19 pandemic. Regarding the third function, for gaining optimal benefit from the CWA, app users must allow the sharing of sensitive data (e.g., time and position data) with the services of the CWA. Therefore, it is of importance that users of the CWA are informed about how such data as well as information will be stored and handled (e.g., misuse of data is prevented). These three tasks were recapitulated using retrospective think aloud (RTA) to identify assessment strategies applied as additional enrichment to recorded eye movements. Afterwards, a set of usability questions (i.e., UQ; based on the System Usability Scale [18]) concerning the CWA was used in the study:

UQ 1: I feel confident using this contact tracing app.

UQ 2: I find this contact tracing app understandable.

UQ 3: I think that my data is well protected in this contact tracing app.

UQ 4: I can benefit from this contact tracing app in the control of the COVID-19 pandemic.

UQ 5: The overall population can benefit from this contact tracing app in the control of the COVID-19 pandemic.

UQ 6: I would recommend this contact tracing app.

The questions were rated on a 4-point Likert scale: strongly disagree (i.e., 1) to strongly agree (i.e., 4). Additionally, an optional open-worded question about missing or desirable functions in the CWA was asked in the study.

D. Study Design

In general, the study at hand was conducted according to the principles of the Declaration of Helsinki. Due to the COVID-19 pandemic, the study was conducted in a prepared lab (e.g., room size of 25 m^2, sufficient ventilation of the room) at Ulm University and the University of Würzburg respectively. Moreover, a dedicated study procedure was carried out in compliance with mandatory COVID-19 hygiene regulations (e.g., permanent wearing of a mouth-nose protection, disinfection of all used instruments before and after each participant). At each study session, only one participant was evaluated and a session took about 20 minutes for completion. The procedure was as follows: In compliance with the hygiene regulations and after a standardized introduction
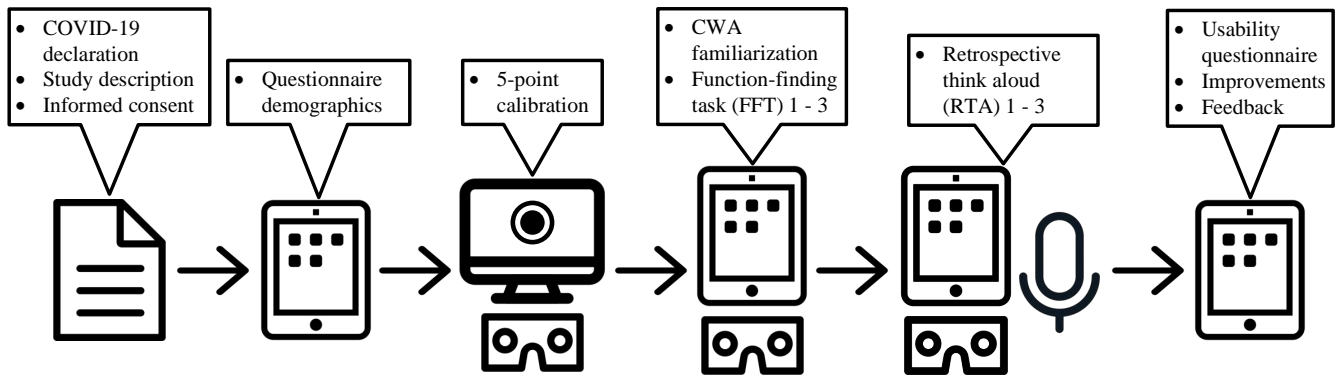
Fig. 1. Study Design

of the participant, she or he needed to sign a COVID-19 declaration (i.e., statement that one is not knowingly infected with COVID-19, is showing no COVID-19 typical symptoms, and has had no contact with known COVID-19 infected persons in the last 14 days). Afterwards, the study procedure was explained and an informed consent was obtained. The participant was then given a tablet, with which demographic data (e.g., gender, age) and information about the CWA and other prominent COVID-19 tracing apps (e.g., CWA, luca App, Stopp Corona) installed on the own smartphone were collected. Participants have been then requested to put on the eye tracking headset (see Sect. II-F) and the device was calibrated on an external monitor with a 5-point screen-based calibration. After completing these mandatory steps, the participant opened the CWA on a prepared tablet. Note that the whole navigation on the tablet was performed with a stylus pen. Following this, she or he got the instruction to look at the CWA and to familiarize with the different functions of the app. The participant could take as much time as needed for this and was instructed to return to the main screen upon completion. Successively, the participant was instructed about three tasks (i.e., FFT; see Sect. II-C), in which a specific function should be found in the CWA, starting from the main screen of the app. During the search, the eye movements were recorded. When the function was found by the participant, the stylus pen was pointed on it, while, at the same time, the command word "Here" was stated. Then, the eye movement recording was stopped and the participant returned to the main screen for the next FFT. After the completion of all three FFTs, a participant needed to recapitulate the applied assessment strategies for finding the respective function. For this purpose, the microphone was switched on and the participant replayed the three FFTs. For each FFT, the participant verbally reflected how she or he found the requested function. In doing so, the participant stated whether certain design aspects (e.g., colors) attracted and assisted during the search for respective function. Here-after, the participant took off the eye tracking headset and answered a set of questions regarding the usability of the CWA on a tablet (i.e., UQ; see Sect. II-C). In addition, she or he could indicate, which functions were missing or desirable

in the CWA. Finally, the study ended after the opportunity to leave feedback. Figure 1 illustrates the design used in the presented study.

E. Measures

The following study measures were considered in the evaluation of the usability of the CWA:

- Duration: The duration (i.e., time spent with CWA) was measured during CWA familiarization and function-finding task 1 - 3 (i.e., FFT; see Sect. II-C). The measured duration for the FFTs allowed us, for example, to draw conclusions about a potential correlation between prior experience with the CWA or related apps and finding the respective function.
- Number of fixations: Fixations constitute eye movements of very low velocity (i.e., eyes remain still on a specific place in a stimulus over a period of time). Moreover, fixation sequence analyses (i.e., sequential analysis of attentional content reading) enabled the identification of conducive or distracting elements (i.e., graphic or textual) in the CWA.
- Keywords: From the RTA, the applied visualized assessment strategies (i.e., eye movements) during the FFTs could be described in more detail. More specifically, concise keywords (e.g., color) were identified from the RTA voice recordings, which related to the eye movements (e.g., gaze attracting elements) and delineated the applied strategies of the participants for finding respective function.
- Usability questions: The obtained answers from the set of usability questions (i.e., UQ; see Sect. II-C) gave us a first impression about the subjective assessment of the usability and usefulness of the CWA during the COVID-19 pandemic.

F. Instrumentation

COVID-19 declaration and informed consent were obtained using pen-and-paper based report forms (i.e., questionnaires). Demographic data, information about COVID-19 tracing apps, questions related to the usability as well as usefulness of the CWA, and feedback were collected using questionnaires in Google Forms on a Samsung Galaxy
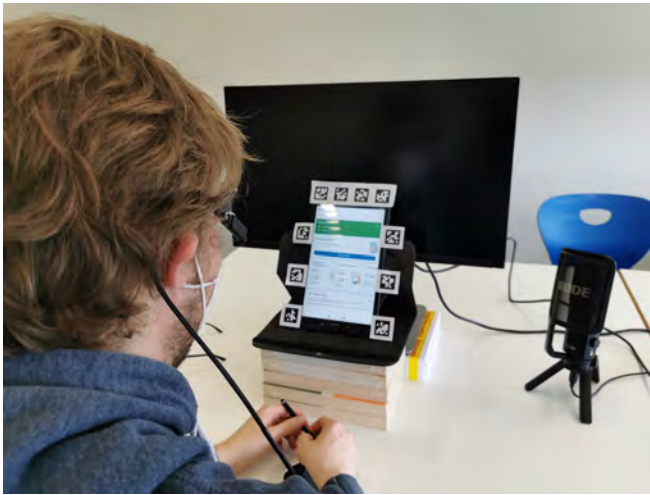
Fig. 2. Setup of the Study Setting

Tab A 10.1. Eye movements were recorded with the Pupil Labs Core headset (200 Hz; 192x192 px), relying on 3D pupil detection [19]. The field of view was recorded with a specific world camera (60 Hz; 1280x720 px) mounted on the headset. The headset was connected to a laptop running with Windows 10, Intel Core i7 8565 U, and 16 GB Ram. Settings for eye tracking were defined in Pupil Capture 3.1.16. For calibration, a 5-point calibration was performed on an external monitor (Dell U2515H, 60Hz; 2560x1440 px; 123.42 PPI; width = 553 mm, height 311 mm). The monitor was positioned 105 cm in front of the participant. Eye tracking data was analyzed and visualized with the Pupil Player 3.1.16. The CWA 1.15.1 was used on a prepared tablet (i.e., Samsung Galaxy Tab A 10.1). In more detail, the edges of the tablet were marked with QR codes in order to isolate the eye movements made within the tablet (i.e., Area of Interest (AOI)) from the ones outside the tablet. The tablet was positioned 100 cm in front of the participant and right in front of the monitor used for calibration. Finally, a Rode NTUSB microphone was used for the RTA. Figure 2 presents the setup of the study setting.

III. RESULTS

In general, Fig. 3 shows the main screen of the CWA with highlighted areas (i.e., red), which represent the position of the three functions (i.e., FFT; ① infection risk, ② COVID-19 test result submission, and ③ sub context menu for privacy policy). Whereas the first two functions (i.e., FFT 1 and 2; see Sect. II-C) are placed on the main screen, it must be navigated through a sub context menu for the third function (i.e., FFT 3). This menu contained, aside the privacy policy, other items with further information (e.g., conditions of use, imprint). From the RTA, gaze-catching elements (i.e., blue; terms and figures) on the main screen ④ *term risk*, ⑤ *term test*, and ⑥ *figure with smartphone* are shown in the figure.

1) Eye Tracking: Table II presents the eye tracking results of the (1) CWA familiarization phase (i.e., CWA Fam) as well as of (2) each function-finding task (i.e., FFT) 1 - 3 (see
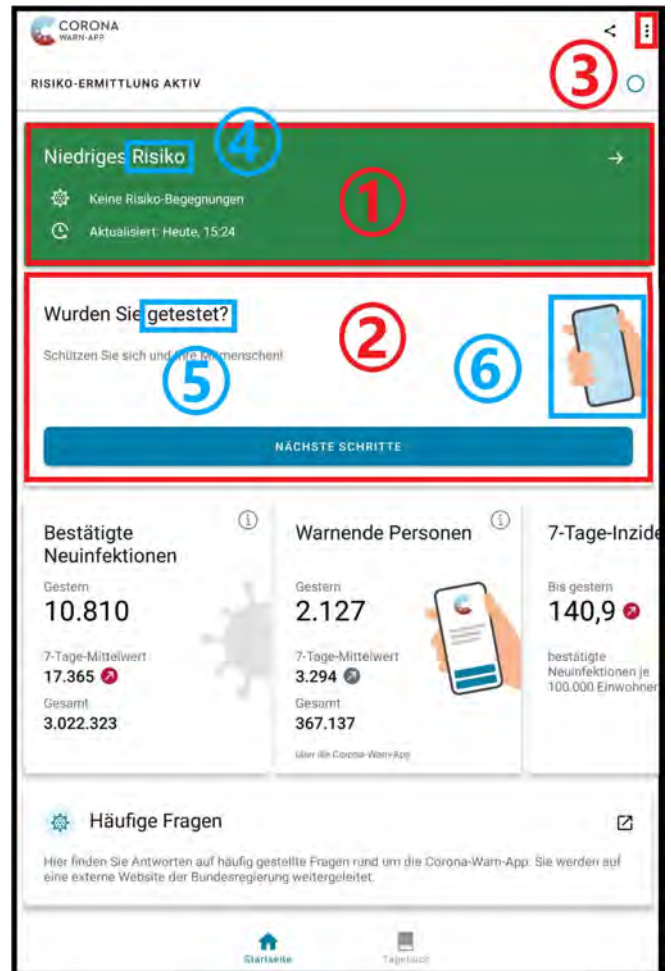

Fig. 3. Main screen of the CWA with highlighted relevant functions (i.e., red) and gaze-catching elements (i.e., blue): ① infection risk, ② COVID-19 test result submission, ③ sub context menu leading to privacy policy, ④ term risk, ⑤ term test, and ⑥ smartphone figure

Sect. II-D). For each participant (i.e., ID) and the entirety (i.e., All) thereof, mean (M) and standard deviation (SD) for the duration (Dur, in seconds (s)) needed and the number of fixations (i.e., No Fix) are shown in the table.

In general, the participants reflected a mean duration of 93.65 (71.13) s (with max = 251.02 s and min = 12.23 s) in CWA familiarization with a mean number of 453.53 (363.75) fixations (with max = 1345.00 fixations and min = 63.00 fixations). For FFT 1, a mean duration of 2.78 (2.98) s (with max = 11.58 s and min = 0.64 s) and a mean number of 13.13 (12.94) fixations (with max = 57.00 fixations and min = 4.00 fixations) were needed. For FFT 2, a mean duration of 1.75 (0.98) s (with max = 3.94 s and min = 0.62 s) and a mean number of 9.93 (5.38) fixations (with max = 23.00 fixations and min = 4.00 fixations) were needed. For FFT 3, a mean duration of 35.87 (38.04) s (with max = 128.45 s and min = 6.12 s) and a mean number of 148.00 (155.79) fixations (with max = 642.00 fixations and min = 24.00 fixations) were needed. During the familiarization phase with the CWA, most of the participants spent a great time on the main screen and only interacted with the functions thereon. Further, not

ID	CWA Fam		FFT 1		FFT 2		FFT 3	
	Dur (s)	No Fix	Dur (s)	No Fix	Dur (s)	No Fix	Dur (s)	No Fix
1	182.97	990	1.22	9	2.20	13	42.01	193
2	48.98	266	2.14	13	2.75	16	15.67	83
3	64.41	339	2.20	13	3.94	23	125.90	642
4	89.08	389	1.50	6	2.64	17	57.26	173
5	30.44	132	1.94	6	1.21	4	20.07	83
6	179.09	681	1.67	8	1.13	6	21.47	98
7	251.02	1345	1.19	4	1.58	9	27.58	120
8	12.23	63	0.64	4	1.56	9	6.12	36
9	88.37	370	8.34	24	3.42	14	128.45	367
10	202.75	1001	11.58	57	1.25	7	10.78	24
11	34.02	174	1.35	9	0.83	6	28.02	144
12	30.67	169	1.12	9	0.62	5	16.26	94
13	73.58	301	4.20	20	1.29	10	14.28	48
14	63.85	308	1.13	5	1.12	4	7.45	32
15	53.45	275	1.51	10	0.74	6	16.68	90
All	93.65	453.53	2.78	13.13	1.75	9.93	35.87	148
(SD)	(71.13)	(363.75)	(2.98)	(12.94)	(0.97)	(5.38)	(38.04)	(155.78)

Note: CWA Fam = Corona-Warn-App familiarization; FFT = function-finding task; Dur = duration (in seconds); No Fix = Number of fixations

much time was spent on reading (i.e., focus on headings), but the eyes kept moving between gaze-catching elements (e.g., colors, figures, buttons). The obtained duration suggests that a first impression about the CWA does not require much time. Hence, the CWA is not bloated with too many functions and focuses its scope to only a few functions. For FFT 1 and 2, the mean duration as well as the number of fixations indicate that these two functions (i.e., infection risk (FFT 1) and positive COVID-19 test result submission (FFT 2); see Sect. II-C) could be found quickly in the CWA. Thereby, it made no difference whether a participant had experience with the CWA or how much time was previously spent for familiarization with the CWA. This is an important indication for an appropriate placement of these two essential functions in the app. Moreover, the duration for FFT 2 was shorter on average than for FFT 1. Further, fewer number of fixations can be seen in FFT 2 compared to FFT 1. A reason could be that after CWA familiarization and the answering of FFT 1, a learning effect has occurred regarding the presentation of the functions within the CWA. Furthermore, regarding recorded eye movements, the gaze of the participants was goal-oriented and only little distracted

Fig. 4. Heat map of a participant for FFT 2

by non-relevant elements. Moreover, several elements on the main screen of the CWA exist, which attracted the gaze of the participants (see Figs. 3 ④, ⑤, and ⑥; RTA Results). For example, Fig. 4 shows a heat map (i.e., magnitude of visual attention (i.e., fixation density)) from a participant in FFT 2. Based on Fig. 3 ②, it can be seen from Fig. 4 that the fixation density is the highest on respective area (red = high fixation density). Regarding FFT 3, however, finding this function (i.e., privacy policy) required more interaction with the CWA, as it can be found in a sub-context menu of the app. Usually, the placement of such information in other apps is often found in corresponding (hidden) context menus on the main screen. Accordingly, the search for this function influenced the eye movements of the participants. Hence, the gaze was spread especially at the beginning of the function-finding task over the main screen, until the more options icon (see Fig. 3 ③) was found. However, as this menu contained other items (e.g., settings) as well, several participants still needed more time to identify the correct menu item.

RTA FFT 1	Color (N = 15), Term risk (N = 10), Main screen (N = 7), Important (N = 5)
RTA FFT 2	Term test (N = 14), Top-Down (N = 9), Figure (N = 5)
RTA FFT 3	Misleading (N = 11), Unapparent (N = 10) General app experience (N = 9)

Note: RTA = retrospective think aloud; FFT = function-finding task

2) RTA Results: The following Table III shows identified keywords obtained from the RTA. In more detail, the table summarizes verbalized assessment strategies for each function-finding task (i.e., FFT) 1 - 3 and are related to recorded eye movements. Therefore, the voice recordings from the participants were evaluated and pertinent keywords were extracted thereof. Thereby, the table shows only those
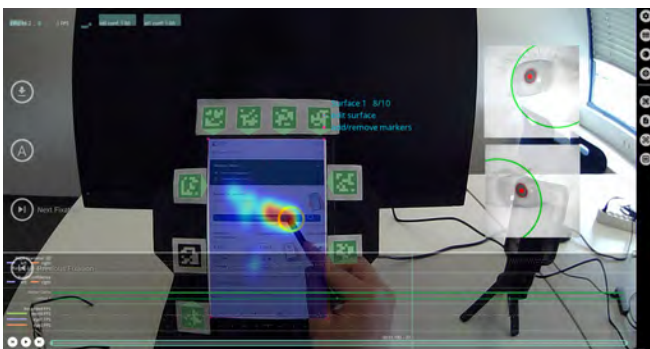
keywords, which occurred with a frequency (N) greater than five from all records. Note that synonyms (e.g., important and relevant) were standardized.

For FFT 1, according Fig. 3 ① and ④, the color and the term risk were gaze-catching elements for finding respective function. Usually, such a central function is placed visibly at the top of an app. The color thereby switches depending on the infection risk: green = low risk, red = increased risk. Hence, the colored banner for the infection risk represents an adequate gaze catcher (i.e., like traffic lights), which is immediately recognizable when opening the CWA. For FFT 2, this function is placed next beneath the first function, when looking at the CWA in a top-down manner (see Fig. 3 ②). From the RTA, this strategy (i.e., top-down approach) was applied by most of the participants. They assumed that this function needs to be close to the first function, as these two functions represent essential functions of the CWA for COVID-19 control. Further, the term test (see Fig. 3 ⑤) represented a supplementary clue and, in addition, the figure of a hand holding a smartphone (see Fig. 3 ⑥) acted as a supportive stimulus in FFT 2. For FFT 3, some participants had difficulties in finding this function. Frequently, this function was confused with another one that is placed on the main screen (i.e., frequently asked questions) of the CWA. The more options icon (see Fig. 3 ③) that leads to the privacy policy could not be recognized directly. Experiences with other apps were helpful, since the more options icon usually represents a sub-context menu with further information. On smaller displays (e.g., smartphone), this function could be even more difficult to find. As this menu contained other items as well, it was not clear to some participants, in which item this information was to be found. Presumably, this function would be found faster if participants had spent more time during the familiarization with the CWA or, in turn, the more options icon is represented more conspicuously.

3) Usability Questionnaire Results: Table IV shows the frequencies of the answers for the usability questions (i.e., UQ) 1 - 6 (see Sect. II-C). The participants indicated that the CWA is easy to use (i.e., UQ 1) and comprehensible (i.e., UQ 2). The duration needed to familiarize oneself with the CWA confirms this observation. Regarding the handling of personal data (i.e., privacy policy, UQ 3), 4 participants stated that they were not sure whether personal data is well-protected in the CWA. Of interest are the results concerning UQs 4 and 5, while participants were at the strife regarding the benefits of the CWA for oneself, in turn, they indicated that the CWA is beneficial for the overall population in the control of the COVID-19 pandemic. This might be an indication that an individual relies more on others, while not using the CWA oneself. A reason could be an ego-centric bias (e.g., myself does not get infected with COVID-19 and, hence, the CWA is not necessary, but others get infected). However, if this attitude is found in large numbers of the population, only a small part would use the CWA. Consequently, the effectiveness of the CWA regarding the control of COVID-19 would be completely lost. Finally, most of the participants would recommend the CWA (i.e., UQ 6).

TABLE IV
RESPONSES IN THE USABILITY QUESTIONNAIRE

	Strong Disagree	Somewhat Disagree	Somewhat Agree	Strong Agree
UQ 1	0	0	6	9
UQ 2	0	1	7	7
UQ 3	0	4	4	7
UQ 4	2	5	6	2
UQ 5	1	2	7	5
UQ 6	0	4	8	3

Note: UQ 1: I feel confident using this contact tracing app.
UQ 2: I find this contact tracing app understandable.
UQ 3: I think that my data is well protected in this contact tracing app.
UQ 4: I can benefit from this contact tracing app in the control of the COVID-19 pandemic.
UQ 5: The overall population can benefit from this contact tracing app in the control of the COVID-19 pandemic.
UQ 6: I would recommend this contact tracing app.

In summary, the implementation and resulting perception of the two essential functions in the CWA (i.e., detected infection risk (FFT 1) and submission of a positive COVID-19 test result (FFT 2)) for the identification of infected individuals and, thus, interrupting infection chains were found to be satisfactory. However, FFT 3 (i.e., privacy policy) revealed that the respective information should be placed more accessible. Further, regarding additional comments from the open-worded question about missing and desirable functions of the CWA, it has been unveiled that information about data handling (e.g., time, position) could be better presented. For example, a detailed and transparent explanation (e.g., short animation) in a prominent visible privacy section about data handling, when using the CWA for the very first time. Since privacy concerns might impede a widespread use of the CWA, this seems to be a crucial aspect in digitization as well. In this context, the results from a survey presented by the authors in [20] indicate that concerns about the privacy policy constitutes a significant factor, which discourage the use of the CWA in large segments of the population. Based on the latter, and the insights obtained from this study, it appears that usability may not be the only primary factor that discourages the population from using the CWA, but rather concerns about privacy policy, which is discussed controversial and discordant in Germany as well as in the Europe Union [21]. Further, the option to present infection numbers at a more fine-grained level (e.g., hometown) is a desirable feature, since several participants stated that the nationwide infection numbers are too abstract. The lack of local availability (i.e., redirection to an external website) of some information in the CWA (e.g., frequently asked questions) was criticized by some participants. Finally, user guidance, especially for aged or inexperienced users, could be therefore addressed (e.g., apparent display of menus) in the CWA.

A. Limiting Factors

The research reported in this paper is subject to several limitations: First, the sample of the study is only limited representative for the target population of the CWA (i.e.,

all mobile devices users in Germany). For example, the age distribution was differing from all mobile device users. As familiarity with mobile devices and app menu trees vary with age, a broader age range of participants must be addressed in future usability studies. Further, the participants were recruited at universities (i.e., Ulm and Würzburg). Second, mobile device fragmentation constitutes another limitation. More specifically, only one tablet (i.e., Samsung Galaxy Tab A 10.1) with a fixed display size was used in the study and other devices (e.g., smartphones with a smaller display) might result in different usability assessments of the CWA (although the general layout is the same on mobile devices). Third, in the meantime, updates already rolled out of the CWA might have addressed potential usability issues. Fourth, the study setup represents another limiting factor. As the participants have always started to search from the main screen of the CWA, the gaze could already focus at the beginning on the functions to be searched for. A randomization in the sequence of the function-finding tasks should be considered as well. Fifth, the current COVID-19 pandemic as well as applicable hygiene regulations (e.g., permanent wearing of a face mask) may have influenced the participants. Finally, the obtained results look promising, however, additional studies are needed either through replication or similar studies to confirm the generalization of the results.

IV. Conclusion and Outlook

In the context of the COMPASS project*, this paper presented exploratory results regarding the usability of the German COVID-19 tracing app Corona-Warn-App (CWA). In particular, in the context of a combined eye tracking and retrospective think aloud (RTA) approach, N = 15 participants evaluated the CWA with respect to usability and usefulness. The obtained results indicate that essential functions (i.e., feedback on infection risk, submission of a positive COVID-19 test result) of the CWA can be quickly found, even without prior app experience. In general, the CWA bears enough quality regarding the control of the COVID-19 pandemic (i.e., identify infected and interrupt infection chains), but several (design) issues exist (e.g., hidden privacy policy) that could be addressed in future updates. Furthermore, as seen in another work presented in [20], usability reasons might not be the only primary issue regarding the timid use of the CWA in the population. Rather, it shows that misleading dissemination strategies or the current European/German privacy policy debate present possible influencing factors, which prevent a widespread use of the CWA among the population. In turn, taking the revealed concerns into account might improve the long-term effectiveness of the CWA in the control of the present COVID-19 pandemic.

While the first exploratory results look promising, the conclusion have to be regarded as preliminary, because of several confronted limitations of the study design (e.g., study population, device fragmentation; see Sect. III-A) that need to be addressed in future work. Therefore, a similar eye tracking and RTA study is currently in preparation, which will compare contact tracing apps from other countries to the CWA regarding usability and usefulness in the context of the COVID-19 pandemic.

References

[1] B. Javed, A. Sarwer, E. B. Soto, and Z.-u.-R. Mashwani, "The coronavirus (covid-19) pandemic's impact on mental health," *The International Journal of Health Planning and Management*, vol. 35, no. 5, pp. 993–996, 2020.

[2] V. Letzel, M. Pozas, and C. Schneider, "Energetic students, stressed parents, and nervous teachers: A comprehensive exploration of inclusive homeschooling during the covid-19 crisis," *Open Education Studies*, vol. 2, no. 1, pp. 159–170, 2020.

[3] U. Frick, M. Tallon, K. Gotthardt, M. Seitz, and K. Rakoczy, "Cultural withdrawal during covid-19 lockdown: Impact in a sample of 828 artists and recipients of highbrow culture in germany." *Psychology of Aesthetics, Creativity, and the Arts*, 2021.

[4] A. Remuzzi and G. Remuzzi, "Covid-19 and italy: what next?" *The Lancet*, vol. 395, no. 10231, pp. 1225–1228, 2020.

[5] K. H. Grantz, H. R. Meredith, D. A. Cummings, C. J. E. Metcalf, B. T. Grenfell, J. R. Giles, S. Mehta, S. Solomon, A. Labrique, N. Kishore *et al.*, "The use of mobile phone data to inform analysis of covid-19 pandemic epidemiology," *Nature Communications*, vol. 11, no. 1, pp. 1–8, 2020.

[6] S. Whitelaw, M. A. Mamas, E. Topol, and H. G. Van Spall, "Applications of digital technology in covid-19 pandemic planning and response," *The Lancet Digital Health*, 2020.

[7] L. C. Ming, N. Untong *et al.*, "Mobile health apps on covid-19 launched in the early days of the pandemic: content analysis and review," *JMIR mHealth and uHealth*, vol. 8, no. 9, p. e19796, 2020.

[8] Robert Koch Institut, "Interrupt chains of infection digitally with the Corona-Warn-App," 2021, https://www.rki.de/EN/Content/infections/epidemiology/outbreaks/COVID-19/CWA/CWA.html, last accessed on 03/05/2021.

[9] J. Wang, P. Antonenko, M. Celepkolu, Y. Jimenez, E. Fieldman, and A. Fieldman, "Exploring relationships between eye tracking and traditional usability testing data," *International Journal of Human–Computer Interaction*, vol. 35, no. 6, pp. 483–494, 2019.

[10] O. Asan and Y. Yang, "Using eye trackers for usability evaluation of health information technology: a systematic literature review," *JMIR Human Factors*, vol. 2, no. 1, p. e4062, 2015.

[11] T. Alshammari, O. Alhadreti, and P. Mayhew, "When to ask participants to think aloud: A comparative study of concurrent and retrospective think-aloud methods," *International Journal of Human Computer Interaction*, vol. 6, no. 3, pp. 48–64, 2015.

[12] V. Kumar, "The think aloud method: Some concerns addressed," *Journal of Modern Languages*, vol. 15, no. 1, pp. 13–25, 2017.

[13] H. Cho, D. Powell, A. Pichon, L. M. Kuhns, R. Garofalo, and R. Schnall, "Eye-tracking retrospective think-aloud as a novel approach for a usability evaluation," *International Journal of Medical Informatics*, vol. 129, pp. 366–373, 2019.

[14] O. Alhadreti, F. Elbabour, and P. Mayhew, "Eye tracking in retrospective think-aloud usability testing: Is there added value?" *Journal of Usability Studies*, vol. 12, no. 3, pp. 95–110, 2017.

[15] O. Alhadreti and P. Mayhew, "Rethinking thinking aloud: A comparison of three think-aloud protocols," in *Conference on Human Factors in Computing Systems 2018, CHI'18*, 2018, pp. 1–12.

[16] L. Helle, "Prospects and pitfalls in combining eye-tracking data and verbal reports." *Frontline Learning Research*, vol. 5, pp. 1–12, 2017.

[17] C. Vogel, R. Pryss, J. Schobel, W. Schlee, and B. Felix, "Developing apps for researching the covid-19 pandemic with the trackyourhealth platform," in *34th Symposium on Computer-based Medical Systems 2021, CBMS'21*, 2021, accepted for publication.

[18] J. R. Lewis, "The system usability scale: past, present, and future," *International Journal of Human–Computer Interaction*, vol. 34, no. 7, pp. 577–590, 2018.

[19] Pupil Labs, "Pupil Core," 2021, https://pupil-labs.com/products/core/, last accessed on 03/05/2021.

[20] K. T. Horstmann, S. Buecker, J. Krasko, S. Kritzler, and S. Terwiel, "Who does or does not use the 'corona-warn-app' and why?" *European Journal of Public Health*, vol. 31, no. 1, pp. 49–51, 2021.

[21] B. Custers, A. M. Sears, F. Dechesne, I. Georgieva, T. Tani, and S. van der Hof, *EU personal data protection in policy and practice*. Springer, 2019, vol. 1.