Int. J. Human-Computer Studies 97 (2017) 1–22

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Designing emotional support messages tailored to stressors

Peter Kindness [a], Judith Masthoff [b,*], Chris Mellish [a]

[a] dot.rural, University of Aberdeen, UK
[b] Computing Science Department, University of Aberdeen, UK

ARTICLE INFO

Article history:
Received 7 July 2015
Received in revised form
11 July 2016
Accepted 29 July 2016
Available online 1 August 2016

Keywords:
Stress
Emotional support
Adaptation

ABSTRACT

Although computers could offer emotional support as well as task support when aiding a user for a complex task, there is little current understanding of how they might do this. Moreover existing demonstrations of emotional support, though promising, only cover a small number of types of support and investigate a limited number of algorithms designed by hand. In this paper, we present an empirical investigation that starts from first principles, determining different categories of stressors for which emotional support might be useful, different categories of emotional support utterances and promising algorithms for deciding the content and form of textual emotional support messages according to the stressors present. At each stage, the results are validated through empirical experiments with human participants who, for instance, are required to place statements into categories, evaluate possible support messages in different imagined situations and compose their own emotional support from options offered. This development methodology allows us to avoid potentially challenging ethical issues in presenting people with stressful situations. Although our algorithms are attempting to choose emotional support based on the general, "naive" competence of human speakers, we use as a running example situations that can arise when attending a medical emergency and awaiting expert help.

1. Introduction

Computer systems nowadays support humans in many different tasks. Their superior abilities in searching, book-keeping and summarisation render them irreplaceable in many complex situations. However computer systems in general only address the informational needs of a human being. For a stressful task, a human supporter will not only provide information but also attempt to alleviate the undesirable emotions being experienced by the task performer. This is a feature that is only beginning to be addressed in computer systems, typically in the context of virtual agents. Virtual agents are computer-generated virtual characters that interact intelligently with users typically taking on roles that normally performed by humans such as coaches, tutors or customer representatives. By *emotional support* we refer to communications from a supporter to a task performer that do not provide concrete help with the details of the task but attempt to address the emotions that are being invoked by the task. We focus on emotions arising from the stressful nature of the task. For instance, a supporter might reassure "Don't worry", show empathy "I understand that you are feeling frustrated", praise "you are doing a great job" or encourage "you can do this". Human beings seem to be remarkably successful at giving emotional support. At least, we trust them to give emotional support in key situations, often with very little training. Moreover, they are able to adapt their support to the type of situation being experienced, which is important as support provided in the wrong context can have a detrimental effect (cf. Lehman et al. (1986), as reported in Cutrona and Suhr (1992)). This paper is about initial attempts to produce a computer algorithm able to capture some aspects of this human behaviour, in particular able to adapt emotional support to different stressors.

In order to ask the question "what emotional support should be given in this situation?", we start by addressing some more fundamental questions:

1. What kinds of stressful situations are there?
2. What kinds of emotional support are there?
3. How is it possible to elicit examples of human emotional support for multiple types of situations in an ethically appropriate way?

The task is then that of modelling the human behaviour as a computer algorithm[1] in particular deciding:

* Corresponding author.
E-mail address: j.masthoff@abdn.ac.uk (J. Masthoff).

[1] One can argue about whether mimicking human behaviour is the right approach. In many fields of Artificial Intelligence, the focus has traditionally been on emulating humans, for example when modeling emotions (Gratch and Marsella, 2001) or generating natural language (see van Deemter (2016) for a discussion on this). Given emotional support in computers is still in its infancy and humans interact with computers in a social way (Nash et al., 1994), modelling human behaviour seems an appropriate first step. However, this will be combined with measuring the effectiveness of emotional support from the receiver's point of view, similarly to the work in Paraboni et al. (2007) on measuring effects of natural language utterances on hearers.

http://dx.doi.org/10.1016/j.ijhcs.2016.07.010

1. Which type(s) of statements should be included for each type of situation?
2. How should the material in the emotional support be ordered?

The work of this paper was inspired by the MIME project (MIME, 2013), which investigated the development of a computer aid for Community First Responders (CFRs) attending medical emergencies. CFRs are volunteers with limited medical training who attend medical emergencies, particularly in remote and rural areas, while an ambulance is en route. The computer aid enabled CFRs to measure and monitor the key medical parameters of the casualty, enter information about their observations and actions, and generated a handover report for the ambulance personnel when they arrived. Although the task of a CFR is known to be stressful in a number of ways, the MIME system only addressed the provision of factual information to the CFRs. Inspired by this, the research described in this paper asked the question "what sort of emotional support might a computer provide to people experiencing the kinds of stressors CFRs experience?".

Following a review of related work in Section 2, in Section 3 we produce and validate textual scenarios depicting individual stressors, crowd source a corpus of emotional support statements, and reliably categorise these statements into emotional support categories. In Section 4, we use the statements and scenarios to investigate what emotional support people offer to other people experiencing different stressors. Based on this, we develop three emotional support algorithms and evaluate these, leading to a refined algorithm and a further evaluation. An overview of this process is illustrated in Fig. 1. Section 5 concludes the paper and provides indications for future work.

2. Background and related work

2.1. Types of stress

We assume in this paper that emotional support to a person is relevant when there is one or more *stressor* in the environment. A stressor is here just a cause of *stress*, where we use Selye's definition of stress as "the non specific (that is, common) result of any demand upon the body [..] be it a mental or somatic demand for survival and the accomplishments of our aims" (Selye, 1956). The APA categorizes stress into the following three categories (APA, 2013): (1) *acute stress* occurs over short durations of time and comes from pressures from the recent past or anticipated near future, (2) *episodic stress* is when an individual experiences one episode of acute stress which is then followed shortly and frequently by another episode, (3) *chronic stress* is a long term experience which is continuous over a long duration of time, such as months and years. We only consider emotional support for acute stress.

In previous work, to identify different stressors for CFRs, we started from the NASA-TLX (Hart, 2006), a multi-dimensional subjective scale for measuring task workload developed by the American National Aeronautics and Space Administration. The NASA-TLX had been used to record participants' perception of task demands, such as those experienced by firefighters (Young et al., 2013) and had been used in various domains due to its high reported levels of internal validity (Xiao et al., 2005) and simplicity. The NASA-TLX measures six workload facets: mental demand, temporal demand, physical demand, frustration, effort, and performance. We excluded effort and performance, as it was assumed that CFRs would always give their maximum effort and that to measure performance would be inappropriate. Based on focus groups with CFRs, we added the emotional aspects of what CFRs may experience while fulfilling their duties, namely 'worries about whether you were doing the right thing', 'irritation by external factors such as interruptions', 'feeling alone and isolated', and 'being upset about (anticipated) outcome'. The resulting set was used to investigate which stressors CFRs experience the most and during what moment of the call-outs (see Kindness et al., 2014 for the results of the survey). The current work takes this list of stressors as a starting point, and considers mental demand, temporal demand, physical demand, emotional demand, frustration, interruption and isolation (see Fig. 2 for informal glosses of these terms).

2.2. Emotional support

Emotional support is an important social construct and a skill which people develop from an early age. However, it is not entirely straightforward and without complications. The same emotional support may not be equally suitable in different situations. Some people, such as counselors, might be considered as being better at providing appropriate emotional support than others. As different stressors are present in different stressful events, it seems plausible that supportive messages should also vary, with some being considered better suited and more beneficial for one stressor than another (Cutrona and Suhr, 1992). Emotionally supportive actions are those aimed at increasing positive emotional states in conjunction with helping people overcome negative emotional states (Burleson and Kunkel, 1996). Emotional support can be considered as a synonym for comforting support. Comforting support can be defined as being "messages having the intended function of alleviating or lessening emotional distress…" (Burleson, 1985). We applied the same definition to emotional support and when the term *emotional support* is used in this research it refers to this definition.

When received, emotional support has been shown to impact a person's mental and physical wellbeing as well as reduce negative affect such as stress (Burleson and Kunkel, 1996; Meyer and Turner, 2002; Goldsmith, 2004). Even the perception that emotional support is available to a person has been shown to be beneficial and can have important implications for stressors such as loneliness (Pierce et al., 1991).

Furthermore, as well as reducing stress, appropriate and sensitive support also strengthens the bond between the support provider and the receiver (Burleson and Kunkel, 1996). This could be potentially important in strengthening the bond between a user and a virtual agent. However, Burleson and Kunkel (1996) also warn that inappropriate support can exacerbate the recipients' stress and have detrimental effects on the relationship between provider and receiver. This is regardless of how good the intentions of the support provider are. This again promotes the need for further investigation into the different types of emotional support that a virtual agent should provide in different stressful circumstances.

The benefits of emotional support alongside its potential pitfalls accentuate the need for further exploration into the differing types of emotional support which could be provided by a system, such as a virtual agent, for different stressors. As discussed in more depth in Section 3, we will base our emotional support categories on those proposed by Dennis et al. (2013), namely Emotional Reflection, Praise, Emotional Advice, Reassurance, and Directed Action.

There is at present no literature on what types of emotional support would be most appropriate for the stressors investigated in this paper. In terms of the emotional support that embodied agents have provided in numerous studies to participants experiencing stress, there remains little to no distinction between the types of emotional support that should be given for different stressors. Researchers working on computer-generated emotional

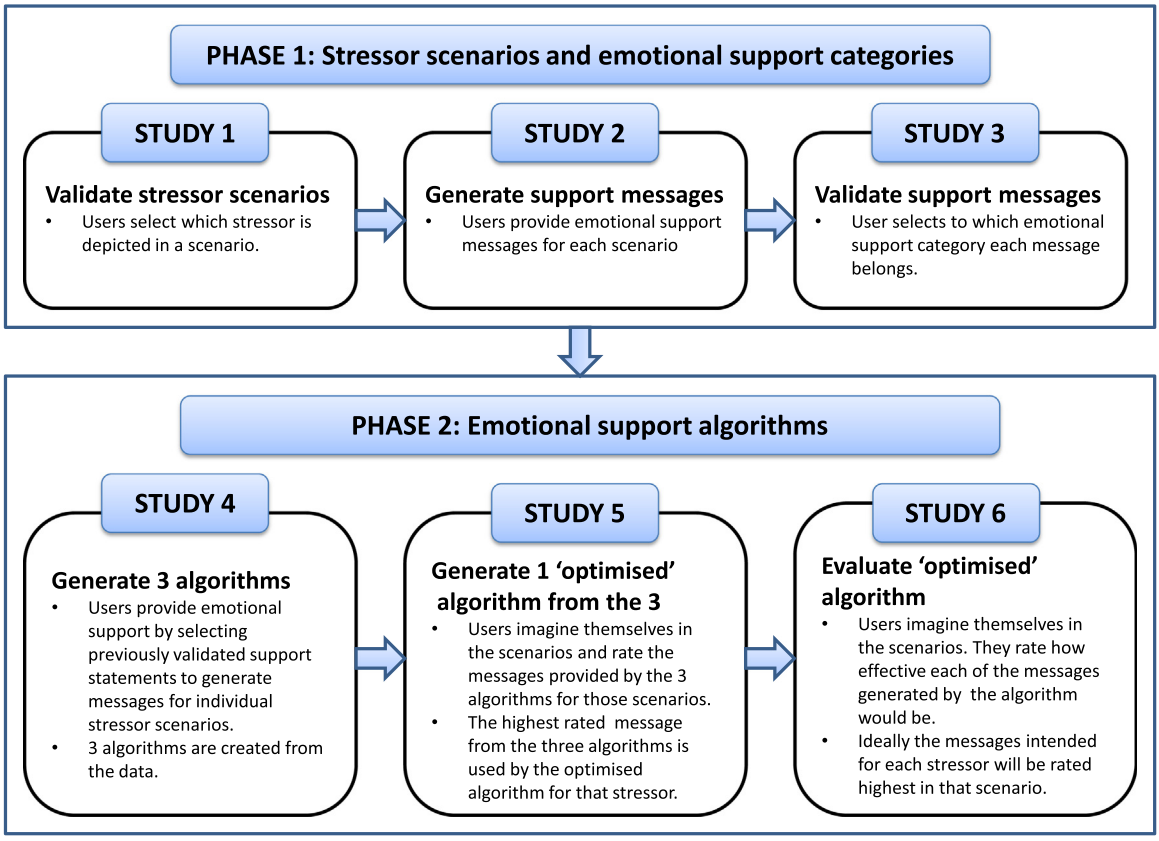

Fig. 1. An overview of the paper's studies. Phase 1 is reported in Section 3 and Phase 2 in Section 4.

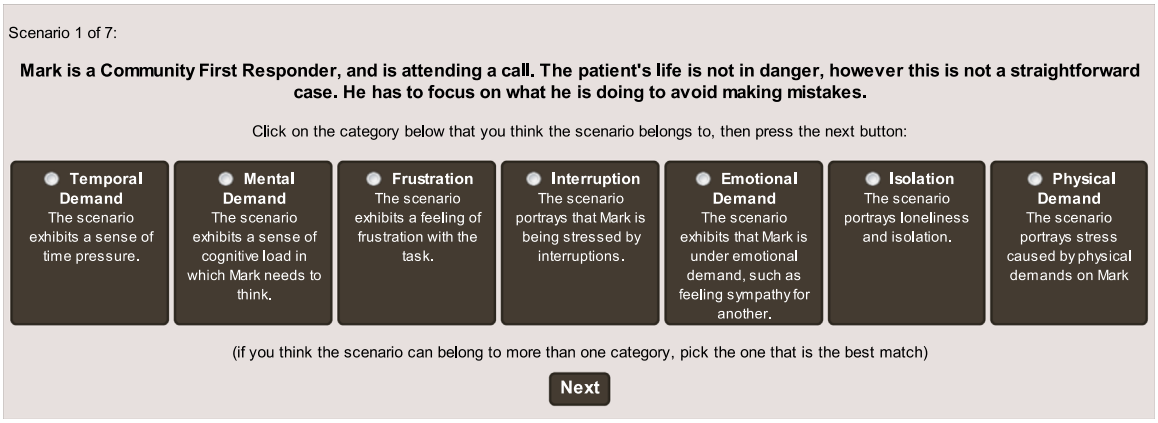

Fig. 2. A screen capture of the 'Validating Scenarios that Exhibit Specific stressors' study.

support have tended to focus only on the stressor frustration. The emotional support provided in these studies (such as in Prendinger and Ishizuka (2005)) has also often considered support to be more of a '*one-size fits all*' approach, with it being considered emotional support as long as it contains either empathy and or encouragement/praise. We will explore the related work more in the next subsection.

2.3. *Agents providing emotional support to alleviate stress*

Several researchers have designed agents to alleviate negative moods experienced by users in stressful situations. In this section, we explore which emotional support algorithms these agents used. We focus here on agents specifically targetting a stressor, that have been implemented, and the effectiveness of which has been evaluated (more work on computer-generated emotional support will be referred to when discussing emotional support categories in more detail in Section 3.4). Table 1 provides a summary.

Klein et al. (2002) used a frustrating computer game and compared the user's frustration levels between a condition that allowed the user to vent their emotions and a condition that ignored the user's emotions completely. The system provided affective support adapted from the participants' input to a questionnaire. This ranged from, "*It sounds like you were not frustrated in the least playing this game. Is that about right?*" when the participant was least stressed, to "*Good heavens! It sounds like [you] felt unbelievably frustrated and completely fed up playing this game. Is that about right?*" for when they were most stressed. However, it is not clear from the paper which categories of emotional support were expressed. Although these statements were not validated or categorized into emotional support categories, we believe

Table 1
Examples of responses that affective embodied agents have applied in stressful situations in existing related work and which of our defined emotional support categories they belong to.

Author	Stressor	Actions	Example response	Similar to
Klein et al. (2002)	Frustration	Affect-support	"It sounds like you were not frustrated in the least playing this game. Is that about right?", "Good heavens! It sounds like [you] felt unbelievably frustrated and completely fed up playing this game. Is that about right?"	Emotional Reflection
Hone (2006)	Frustration	Affect-support	"Oh dear, it seems like you're feeling frustrated."	Emotional Reflection
Prendinger et al. (2003)	Frustration	Show Empathy	"I apologize that there was a delay in posing the question"	N/A
Prendinger and Ishizuka (2005)	Frustration	Show Empathy	The agent displays concern for a user who is aroused and has a negatively valenced emotion, e.g., by saying "I am sorry that you seem to feel a bit bad about that question."	Emotional Reflection
		Encourage	"You appear calm and do not have to worry. Keep going!"	Reassurance
		Ignore	"Let us go on to the next question."	N/A
		Congratulate	"Well done!", "Good job!", "You said the right thing", …	Praise
Baylor et al. (2005)	Frustration	Show Empathy	"It must have been very frustrating trying to finish the survey with the problem you were experiencing. I sympathise with how you feel. […]"	Emotional Reflection
		Apologize	"I'm really sorry that this problem happened to you. I know that the problem could have been avoided on our part, and it was not your fault […]"	N/A

statements of this sentiment could be mapped to our emotional support category *Emotional Reflection*, which aims to acknowledge how the recipient is feeling. Klein et al.'s study showed that the affect-support agent could reduce negative feelings after the users had been frustrated by network delays.

The studies by Hone (2006), which were a partial replication of Klein et al.'s study, used the same method of delivering affective support as Klein et al. did (though using a virtual agent rather than just text messages). Although Hone stated that "participants interacted with an affective agent with similar behaviour to that in [the study by] Klein et al. (2002)", we do not know if the same emotional support messages where applied. Again, as the statements which Hone applied were not categorized, we considered them to belong to the *Emotional Reflection* support category, similar to those applied in the study by Klein et al.. Hone's studies also found that affect-support agents reduce negative feelings after users have been frustrated.

In Prendinger et al. (2003) and a number of similar studies, the emotional feedback that was provided to the user was given promptly after the user experienced frustration and this feedback was also delivered via a lifelike character and not just via text. The affective agent described in the study conveyed that it was either 'happy for' or 'sad for' the participant and expressed emotion through the use of body gesture and text. The agent would either smile to express happiness or hang its shoulders to express sorriness. To express empathy to the participant (who had been subjected to system delays), the agent would utter "*I apologize that there was a delay in posing the question*.". However, unlike in the aforementioned studies, we do not categorize this emotional support as belonging to any of our categories defined below. Although the statement is given in recognition that there was a delay which could have been a source of frustration to the participant and therefore could be categorized as *Emotional Reflection*, it does not explicitly portray this and as such could either be categorized as commiseration or as simply stating a fact. Klein et al.'s study showed that the empathic agent reduced participants' stress level (as measured through skin conductance).

In a further study by Prendinger and Ishizuka (2005), the emotional support provided by the affective agent is more clearly defined. The different categories of support which were used by the agent in this study are shown in Table 1. From the categories of support shown in Table 1, it is clear that there are parallels in our defined emotional support categories and those that were applied by the agent in Prendinger and Ishizuka's study. Prendinger and Ishizuka's agent used four support categories, three of which can be compared to our five defined emotional support categories. 'Show empathy' can be compared to our emotional support category of *Emotional Reflection*, where a sense of understanding for the user's emotions is conveyed. 'Encouragement' can be compared to our emotional support category of *Reassurance*, where the user is reassured about their abilities. And 'Congratulate' can be compared to our emotional support category of *Praise*, where the user's abilities are positively reinforced. The 'Ignore' response was a category which lacks any emotional support, in which the user's feeling are not recognized by the agent. We do not have a similar emotional support category to Prendinger and Ishizuka's 'Ignore' response. While their study did not find an overall effect of the emphatic agent, there was some evidence that emphatic feedback helped lower stress levels while the participant listened to interviewer questions (in a virtual job interview setting).

Baylor et al. (2005) studied a virtual agent that reacted to frustration caused by a pop-up error message that provided an obstacle to answering survey questions, comparing between an emphatic agent and an apologetic one. In their work, the agent was part of the system that caused the frustration, hence the potential appropriateness of an apology. We do not have a similar emotional support category to Baylor et al.'s 'Apology' response. Their 'Empathy' response is similar to our *Emotional Reflection* category. Both versions of the agent also encouraged the user to provide feedback, so they could vent their emotions. In contrast to the other studies mentioned, Baylor et al.'s study actually found that both conditions increased user frustration, though users felt the emphatic agent was more believable. It is possible the venting option combined with the type (and amount) of frustration may have caused this, as this frustration was clearly the system's fault (more clearly than the network delay used by Klein et al).

In our studies, described later, we had two additional emotional support categories: *Emotional Advice* (where the person is told how to feel) and *Directed Action* (where the person is told what to do, or the manner in which to do it). Neither of these emotional support categories were applied in the aforementioned studies.

The presented studies have demonstrated that emotional support when provided by virtual agents can affect a person's emotional state. However, these studies have only focused on the stressor of *frustration*, and the source of this frustration has been from the system itself (i.e. the system is to blame). In the domain of pre-hospital care, there may be various different stressors (e.g. from the weather, the seriousness of the patient's condition) and the source of stress would not come from the system providing the support. We have already argued that people provide different emotional support depending on the stressor that

the recipient is experiencing and that when the inappropriate emotional support is provided it may have detrimental effects (Burleson and Kunkel, 1996). This strongly promotes the need for an emotional support generating algorithm which an embodied affective support agent can use to tailor its support accordingly. With this in mind, this paper presents several studies which investigate the development of an emotional support algorithm for the generation of effective and appropriate emotional support for different individual stressors.

3. Phase 1: stressor scenarios and emotional support categories

This section summarizes and extends our previous work (Dennis et al., 2013) that lays the foundation for the remainder of this paper.

3.1. Scenarios that describe individual stressors

To investigate how different types of emotional support are provided for different stressors, we were first required to generate conditions in which participants would be able to empathize with somebody's stressful circumstances. One method for achieving this might be to show participants videos of people experiencing stressful situations or to have participants observe people experience stress in real life. However such an approach would have considerable drawbacks, most notably the ethical issues involved with exposing participants to witnessing potentially distressing scenes and the practicality of observing people experiencing stress life. An alternative and much more viable option would be to devise a series of simple, textual statements in which a person and a stressor are described. We used the CFR role as an inspiration to produce these. These statements would then allow participants to take the place of a virtual agent and provide what they consider to be appropriate support. This approach is referred to as the 'User as Wizard' method (Masthoff, 2006; Paramythis et al., 2010).[2] To simply provide a short sentence stating "John is a CFR and is feeling lonely" however is unlikely to elicit a strong enough feeling within participants so that they would empathize with John and provide true emotional support. As a solution to this we developed a set of stories where each describes a scenario involving a CFR and one particular stressor, using seven from nine stressors previously explored via a survey to CFRs (Kindness et al., 2014): temporal demand, mental demand, frustration, interruption, emotional demand, isolation, and physical demand. Glosses of what these terms were intended to mean are shown in Fig. 2 for our later validation study. The stressors 'worries about whether you are doing the right thing' and 'being upset about the (anticipated) outcome' were excluded as it was deemed that these stressors (which indicate how an individual cognitively interprets an event) would be challenging to accurately describe to someone not in the same situation and if described in a scenario are likely to resemble emotional demand. The hand-constructed scenarios and the stressors they intended to describe are presented in Table 2.

3.2. Study 1: validation of scenarios that exhibit specific stressors

We validated that the scenarios only described the stressor for which they were designed and not those intended to be described by the other scenarios.

3.2.1. Method

3.2.1.1. Participants. Participants were recruited through Amazon's Mechanical Turk service (MT, 2013), a crowd-sourcing tool. On Mechanical Turk, adult participants (called *workers*) complete small tasks made available by *requesters* and are paid a small sum for completing the task successfully. For this study, participants had to be based in the US and have an acceptance rate of at least 90% (meaning that 90% of the work they have previously completed has been accepted by other requesters and graded as being of good quality). Participants were paid $0.50 upon successful completion of the study. Due to the language based nature of the study, potential participants first had to complete an English fluency test, namely the Cloze Test (Taylor, 1953). Workers who failed the test were excluded.

Thirty participants completed the study with 73% being male and 27% being female. 46.7% of participants were aged between 18 and 25, 46.7% were between 26 and 40 and 7.6% were between 41 and 65. The average duration of time taken by participants to complete the experiment was approximately 2.5 min.

Ethical approval was obtained for this study, and all other studies in this paper, from the University of Aberdeen College of Physical Sciences ethics committee. Participants provided informed consent.

Procedure: Participants were introduced to the stress categories and their definitions (which are also shown in Fig. 2). Next, participants were presented with a scenario (as presented in Table 2) and asked to place it into one of the scenario categories (while still seeing the definitions). All the scenario texts were preceded by a statement saying "[INSERT NAME OF CFR] is a Community First Responder, and is attending a call.". At the start of the study, participants were informed that a CFR was a first aider. It was deemed unnecessary at this stage to inform participants fully of the role of a CFR. A screen capture of the study is shown in Fig. 2. This process was repeated for each scenario, and the order of the scenarios was randomized.

Validation measure: The *Free-Marginal Kappa* (Randolph, 2005) was used as a metric for establishing how well categorized the scenarios were and therefore how well they described each unique stressor. The kappa value describes agreement amongst the ratings of participants, with 1 indicating unanimous agreement, 0.7 excellent and 0.4 moderate agreement.[3] To be reliably categorized, the kappa score for the scenario had to be $\geq$0.4.

3.2.2. Results

Table 2 shows the results of validation, as ordered by *Free-Marginal Kappa* (Randolph, 2005). The results presented in Table 2 show most scenarios strongly depict their intended stressor independently, as indicated by the Kappa scores. However, the scenario describing the stressor of *isolation* had only a moderate Kappa (0.4), as some participants identified the stressor as being either *temporal demand* or *frustration*. A classification as temporal demand may have happened because it mentioned a time: "at least half an hour". A classification as frustrating may have happened because it mentioned that the CFR had to wait for an ambulance to arrive. Despite this scenario not being as strongly categorised as the others, the Kappa score was still adequate for it to be deemed as portraying *isolation*.

3.3. Study 2: generation of emotional support statements

The previous study provided a set of validated scenarios each

[2] While the 'User as Wizard' method was inspired by the 'Wizard of Oz' method, it differs in that in the 'User as Wizard' method the participants take the role of the computer.

[3] No universal agreement on the interpretation of kappa values exist, but (rather arbitrary) magnitude guidelines for different types of kappa have been published. We have based our criteria on the Landis and Koch criteria (Landis and Koch, 1977) which regard a kappa of 0.41–0.60 as moderate and 0.61–0.80 as substantial, and the Fleiss' criteria (Fleiss, 1981) which regard a kappa of 0.4–0.75 as fair to good and >0.75 as excellent.

Table 2
Scenarios for classification, the stressors they intended to describe (Stressor), and the stress categorization. All of the scenarios included the prefix "[INSERT NAME OF CFR] is a Community First Responder, and is attending a call." The figures represent the % of participants rating for each category of stressor. Kappa = Free-Marginal Kappa (Randolph, 2005). Bold shows the most popular category.

Scenario	Stressor	MD	TD	PD	FR	IN	ED	IS	Kappa
The weather is awful today and it is really cold and wet. The patient is heavy and needs putting on a stretcher	Physical Demand (PD)	0	0	**100**	0	0	0	0	**1.00**
Andrew needs to work against the clock as he is needed at another accident	Temporal Demand (TD)	0	**100**	0	0	0	0	0	**1.00**
The patient's life is not in danger, however this is not a straightforward case. He has to focus on what he is doing to avoid making mistakes	Mental Demand (MD)	**97**	0	0	3	0	0	0	**0.93**
Charlie is dealing with a casualty, and passers-by keep asking him questions	Interruption (IN)	0	0	0	3	**97**	0	0	**0.93**
Oliver knows what treatment the patient requires, unfortunately he is not legally allowed to provide this treatment as he has not officially finished his training. Instead, he will have to deal with the minor injuries only	Frustration (FR)	0	0	0	**93**	0	7	0	**0.84**
There has been a very serious accident and it is unclear whether the patient will survive. The patient's wife is visibly upset	Emotional Demand (ED)	7	3	0	0	0	**90**	0	**0.77**
The patient is stable but unconscious. There is nobody else around. It will take at least half an hour for the ambulance to get there	Isolation (IS)	0	20	0	13	0	0	**67**	**0.40**

expressing a unique stressor. This allowed us to investigate what type of emotional support people would provide to individuals experiencing each individual stressor. In this study, we were interested in gathering a corpus of emotional support statements which could be used to support CFRs experiencing such stressors. To achieve this, we ran a simple data gathering experiment which asked participants to provide emotional support to the CFR portrayed in each of the seven validated scenarios.

3.3.1. Method

Participants: The study was run on Mechanical Turk, with the same conditions as in Section 3.2.1.1, including the English fluency test. Twenty participants took part and were 60% male and 40% female. 20% of participants were aged between 18 and 25 years, 45% between 26 and 40 and 35% between 41 and 65. The average time to complete the study was approximately 8 min. Participants were paid $0.50.

Procedure: Participants were shown each of the seven scenarios from Table 2 and shown some examples of potential emotional support statements (taken from a different domain). Participants were asked to provide three examples of emotional support for each scenario.[4] The scenarios were presented in a randomized order. Participants were reminded that they were not expected to give any medical advice (decision support), only emotional support. Participants were asked at the end of the study if they had any comments. A screen capture of this study is shown in Fig. 3.

3.3.2. Results

Each participant gave three support statements per scenario, resulting in a total of 420 statements. The statements were processed during a focus group of researchers to filter out statements which were not considered to be emotional support (e.g. "Call 911"). Inappropriate statements were those that were considered to be decision support, unsuitable to a general population (e.g. religious sentiment), or would not be suitable when provided by a virtual agent (e.g. "I will handle the passers-by so you do not need to deal with them"). Of the remaining statements, duplicates and semantically similar statements were removed. This process resulted in 85 unique statements.

[4] We asked for three examples to ensure we would obtain a varied set of statements, considering also that the first statement people think of may not necessarily be the most effective. Response times indicate that participants had no difficulty generating three statements.

3.4. Deciding upon emotional support categories

To decide upon emotional support categories for the generated statements, we used two approaches. Firstly, we ran an open card sorting session with three other members of our research department, where related statements were grouped together into categories. Secondly, we considered emotional support types used in the literature to try to make sense of the groupings resulting from the card sorting.

Table 3 shows emotional support types used in the literature related to our initial categories from the card sorting. We decided that the existing classifications of empathy, active listening and to an extent sympathy portrayed the sense of understanding how a person was feeling. We therefore decided to use the category Emotional Reflection. We used the category Praise as the support types mentioned under Praise in Table 3 all include statements that make people feel better about themselves. Following much discussion, it was decided that there was no clear distinction between Reassurance and Encouragement, with both seeking to motivate a person when things might not be going as well as expected. We therefore decided to merge these categories and use Reassurance. We decided that Advice was too broad, and similarly to Dennis et al. (2013) divided this into Directed Action (informing a person what to do, or the manner in which to do it) and Emotional Advice (telling a person how to feel). Fig. 4 shows the five resulting categories and their definitions. These categories were also applied to motivate learners by Dennis et al. (2013).

We also considered the literature in sociology and psychology to see how well these categories cover existing support strategies. Within sociology, social support researchers tend to distinguish five types of social support, namely informational, emotional, tangible, esteem and social network support (Cutrona and Suhr, 1992). Tangible support (e.g. offers to provide goods and services) and social network support (e.g. the sense of belonging to a group with similar interests) are outside the scope this paper. Our support category Directed Action provides informational support, defined by Cutrona and Suhr as advice, factual input and feedback on actions. Our category Praise provides esteem support, defined by Cutrona and Suhr as expressions of regard for somebody's abilities, skills and intrinsic value. Our category Emotional Reflections provides what Cutrona and Suhr call emotional support, which they define as expressions of caring, empathy, sympathy and concern. Cutrona and Suhr do not explicitly mention Reassurance or Emotional Advice, but these seem to fit best within their overaching concept of nurturant support (efforts to comfort or console), and from its subcategories fit better with emotional support than with esteem or network support. In this paper, we call all our support categories emotional support, as they can aid users to regulate their emotions.

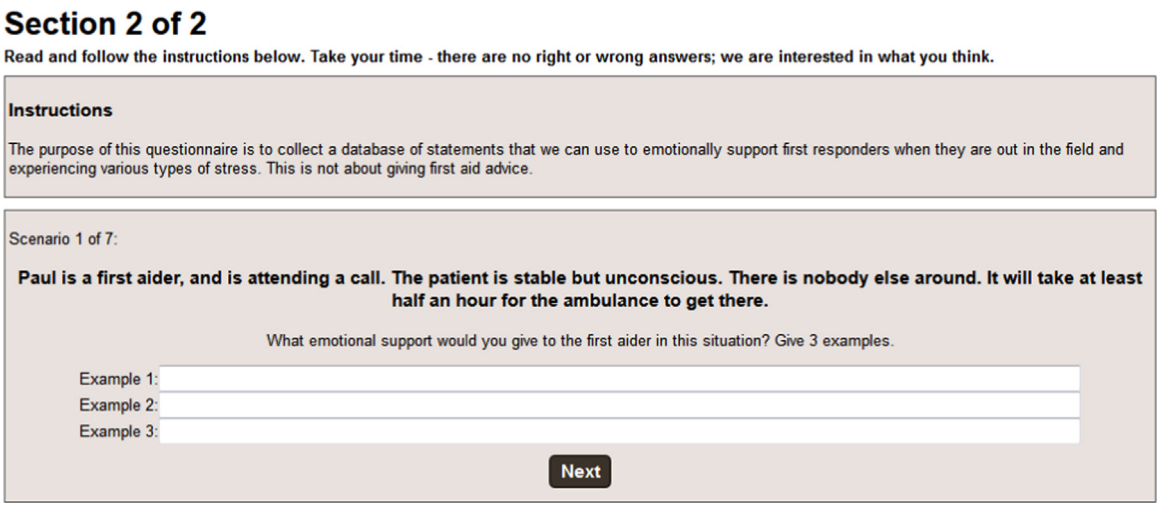

Fig. 3. A screen capture of the 'Generating Support Statements for First Responders' study.

Within psychology, there are two interrelated strands of work that are particularly relevant: (1) coping and (2) emotion regulation. Coping is defined as the "cognitive and behavioural efforts to manage specific external and/or internal demands that are appraised as taxing or exceeding the resources of the person" (Lazarus and Folkman, 1984). Distinctions have been made between problem-focused and emotion-focused coping (Folkman and Lazarus, 1980), between active-cognitive, active-behavioural and avoidance coping (Billings and Moos, 1981), and between approach and avoidance coping (Roth and Cohen, 1986). Many coping strategies exist; we will consider the 14 strategies distinguished by Carver et al. (1989). Table 4 shows three types of coping (combining the categorisations mentioned above), the coping strategies from Carver et al. (1989) that belong to these, and which of our emotional support categories are appropriate to support these strategies. We have not mapped our categories onto the Avoidance coping and Turning to religion strategies. It has been argued that the Avoidance coping strategies are maladaptive (i.e., not having good effects in the long-term) (Zeidner and Endler, 1996) and they are outside our research scope (e.g., we do not intend for the agent to prescribe drugs). More positive results have been found for Turning to religion (McCrae and Costa, 1986), but this strategy is also outside our research scope.

Emotion regulation is defined as the "processes by which individuals influence which emotions they have, when they have them, and how they experience and express these emotions" (Gross, 1998). Gross (2002) distinguishes five types of strategies to regulate emotions: situation selection (approaching or avoiding certain situations to regulate emotions), situation modification (tailoring a situation to modify its emotional impact), attentional deployment (selecting which aspects of the situation to focus on), cognitive change (selecting which meaning to attach to a situation's aspect, e.g. reappraisal), and response modulation

Table 3
Examples of types of Emotional Support extended and adapted from Dennis et al. (2015).

Reference	Empathy	Praise	Advice	Reassurance	Encouragement	Other
Barbee et al. (1993)	Solace		Solve			
Brave et al. (2005), Prendinger and Ishizuka (2004), Paiva et al. (2004)	Empathy					
Burleson and Picard (2007)	Mirroring, Affect					
Cutrona and Russell (1990)	Concern	Love				Interest, Care
Cutrona and Suhr (1992)	Empathy, Concern, Sympathy	Esteem	Advice			Care
Fogg and Nass (1997), Masthoff (1997)		Praise				
Johnson et al. (2004)		Praise, Flattery				
Gilliland (2011)		Praise	Explaining	Reassurance	Encouragement	
Hone (2006), Klein et al. (2002)	Affect support					
Lee (2008)		Flattery				
Nguyen and Masthoff (2009)	Sympathy, Empathy					Perspective
Picard and Klein (2002)	Sympathy, Empathy					Active listening
Prendinger and Ishizuka (2005)	Empathy	Congratulate			Encouragement	
Robison et al. (2010)	Parallel empathy		Reactive empathy			
Rook and Underwood (2000)		Appreciate, Respect		Reassurance	Encouragement	
Dennis et al. (2013)	Emotional reflection	Praise	Directed action, Emotional advice	Reassurance		
Dennis et al. (2015)	Emotional reflection	Praise	Advice, Emotional encouragement	Reassurance		
Smith et al. (2014)	Empathy	Praise, Appreciated	Practical advice, Emotional advice	Blameless, Consolation	Encouragement	Supported, Deserving
Baylor et al. (2005)	Empathy					Apology
van der Zwaan et al. (2012)	Emotional reflection		Advice			Interest

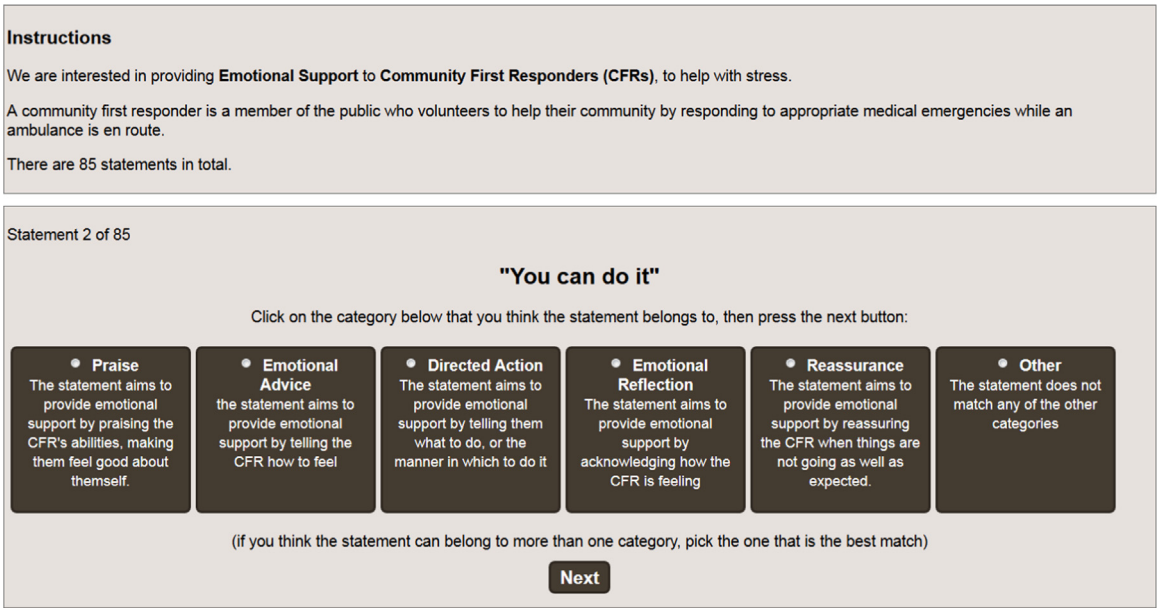

Fig. 4. A screen capture of the 'Categorizing Empathic Support Statements' study.

(influencing emotional response tendencies once they have been elicited, e.g. suppression). Table 4 shows which of our emotional support categories are appropriate to support the five emotion regulation strategy types.

In summary, our emotional support categories cover support types in the literature, cover the social science categories of informational, emotional and esteem support, cover the problem-based and emotion-based coping strategies (with the exception of turning to religion), and cover the emotion-regulation strategies. It may be possible to produce more detailed subcategories in future in particular within Directed Action, to distinguish between Directed Action aimed to support each of the problem-based coping strategies, and/or to support each of the situation modification, situation selection and attentional deployment emotion regulation strategies.

3.5. *Study 3: validation of emotional support statements' categorization*

As outlined in the previous section, a corpus of 85 unique emotional support statements for CFRs was obtained and categorized into a preliminary set of five categories. It was required that these statements were validated as belonging to these categories. To achieve this, we ran another validation experiment similar to the study outlined in Section 3.2.

Table 4
Mapping between coping, emotional support, and emotional regulation.

Coping types (Folkman and Lazarus, 1980; Billings and Moos, 1981; Roth and Cohen, 1986)	**Coping strategies** (Carver et al., 1989)	**Emotional support**	**Emotion regulation strategies** (Gross, 2002)
Approach, Active, Problem-focused	Active coping Planning Suppression of competing activities Restraint coping Seeking social support - instrumental	Directed action	Situation selection, Situation modification, Attentional deployment
Approach, Active, Emotion-focused	Positive reinterpretation and growth Seeking social support-emotional Acceptance Focus on and venting of emotions	Reassurance, Praise Empathy, Emotional advice	Cognitive change Response modulation
Approach, Active, Problem-focused, Emotion-focused	Turning to religion		Cognitive change, Response modulation
Avoidance	Denial Behavioural disengagement Mental disengagement Alcohol-drug disengagement		Situation selection, Situation modification, Attentional deployment, Response modulation

3.5.1. Method

Participants: As before, the validation experiment took the form of an online questionnaire administered on Mechanical Turk, with the same participation criteria as used previously. Forty participants partook in the study and were paid $0.50. Participants were asked to indicate their gender, their age from a range and indicate if they were a health professional. 55% of participants were female and 45% were male. 22.5% were aged between 15 and 25, 45% between 26 and 40, 27.5% between 41 and 65 and 5% were aged over 65. Out of the participants 5% regarded themselves as being healthcare professionals. The average completion time for the experiment was approximately 8 min.

Procedure: Participants were shown an explanation of what being a CFR entailed, the five emotional support categories and their definitions. Next, they were shown each of the 85 statements in turn, and asked to select the category which they felt best fitted the statement. The order in which the statements were presented was randomised. The interface for this study is shown in Fig. 4. Participants could also categorize a statement into an 'Other' category if they felt that it did not fit any of our defined categories. At the end of the study, participants were asked if they had any comments.

3.5.2. Results

Of the 85 statements, 52 were categorized with at least moderate agreement (*Free-Marginal Kappa* ≥0.4). These are reported in Table 5. Of the remaining statements, 29 were weakly categorized (Kappa <0.4), 2 statements had no clear decision (equal rating in more than one category) and 1 was strongly categorized as 'Other'. Out of the five defined emotional support categories, *praise* and *directed action* were the most reliably categorized, with at least one statement in each having complete agreement (Kappa=1) amongst participants. When reviewing the total number of statements, *directed action* had the most statements, with *emotional reflection* having the least. This may indicate that our corpus simply had less examples of *emotional reflection* to begin with. However, this does not necessarily indicate that participants wanted to use less of this type of statement, as there may have been multiple examples of the same statement being used in the corpus,[5] which would only result in one statement required to be validated. Table 5 also shows the top four statements for each category. These were the statements that we took forward into our studies about emotional support generation.

Table 5
The most strongly categorized statements for each category with K ≥ 0.4. CAT= Rated Category, K=Free-Marginal Kappa (Randolph, 2005). DA=Directed Action, EA=Emotional Advice, ER=Emotional Reflection, PR=Praise, RE=Reassurance. Italicised statements were used in the next study.

Statement	CAT	K
Pay attention to detail	DA	1.00
Take your time	DA	0.88
Be efficient	DA	0.88
Try a little harder	DA	0.82
Do not rush	DA	0.82
Take a deep breath	DA	0.73
Stay focused	DA	0.68
Do not be distracted	DA	0.64
Shut everything else out	DA	0.64
Do not lose sight of what you need to do	DA	0.63
Just keep going	DA	0.62
Think like a professional	DA	0.60
Keep a clear head	DA	0.57
Slow down and you will do fine	DA	0.54
Help them as it makes a difference	DA	0.53
Just do what you can	DA	0.50
Give your best effort	DA	0.49
Try your best	DA	0.41
Stay positive	EA	0.69
Be glad that you can help	EA	0.54
Believe in yourself	EA	0.54
Be understanding	EA	0.49
Be compassionate	EA	0.46
Be strong	EA	0.43
I know how you feel	ER	0.77
I understand that this is frustrating	ER	0.72
I know what you are going through	ER	0.63
I know this is hard	ER	0.49
Great work	PR	1.00
Good job	PR	1.00
You are doing a great job	PR	0.88
Good job at keeping your cool	PR	0.82
Thank goodness that you are here and are knowledgeable	PR	0.72
You efforts are appreciated and valued	PR	0.72
You work great against pressure	PR	0.67
You are a pro at this	PR	0.60
You are doing well	PR	0.60
You are capable and competent	PR	0.59
You are handling it well	PR	0.56
You are really helping this situation	PR	0.48
It is going to be fine	RE	0.67
It will be OK	RE	0.67
You will get through this	RE	0.59
This will be over soon	RE	0.58
You can do this	RE	0.50
You can handle this	RE	0.49
You will get there eventually	RE	0.48
I am here for you	RE	0.46
You can do it	RE	0.46
This is complex but you can work through it	RE	0.44
It is not your fault	RE	0.42
We have got plenty of time	RE	0.41
You can only do so much	RE	0.41

4. Phase 2: emotional support algorithms that tailor support to specific stressors

4.1. Study 4: initial emotional support algorithms generation

In this study we investigated which categories of emotional support (if any), people would use when providing support to other people in scenarios exhibiting varying stressors.

4.1.1. Method

Participants: Participants were recruited using Amazon's Mechanical Turk service, similarly as in the previous studies. They were paid $0.50 upon completion of the study. One hundred participants completed the study. 54% were male and 46% female. 26% were aged between 18 and 25, 53% between 26 and 40 and 21% between 41 and 65. The average time for completion was just under 7 min. Participants were also asked to indicate their level of medical expertise. This was to gain an insight into the potential expertise that the participants may have with the scenarios presented and how similar they may be to the medical expertise of CFRs. Only 1% considered their expertise to be advanced, whereas 13% considered it to be intermediate and 53% as novice. Of the participants, 33% considered themselves to have no medical expertise at all.

Materials: The top four most strongly correlated statements from each support category (those with the highest Kappa score) were used (the italicised statements in Table 5). The decision to only select the top four statements was taken for two reasons.

[5] Duplicate and semantically similar statements were removed during an earlier process (focus group as mentioned in Section 3.3.2). Many of the statements that were then validated had been provided by multiple people.

Firstly, in the EMOTIONAL REFLECTION category only four statements were strongly validated. As previously stated, this does not necessarily indicate that EMOTIONAL REFLECTION is a weak emotional support category, only that in the corpus gathering process people may have simply provided very similar statements for this category. Secondly, it was deemed undesirable to overwhelm participants with a vast list of statements from which they would have to choose. If a participant was overwhelmed with a large list of options, it was hypothesized that they may simply pick the first suitable option instead of reviewing the full list and selecting the one they considered to be the most appropriate.

Procedure: Participants were presented with the seven scenarios (as shown in Table 2) which had been validated to describe a unique source of stress. The scenarios were presented in a randomised order. For the purposes of this study, CFRs (who were originally described in the scenarios during their validation) were referred to simply as first aiders.

Participants were asked to provide support to the first aider in the scenario using a set of statements made up from the top four previously validated statements from each emotional support category. These 20 statements were presented in a randomised order. Participants could provide multiple statements of support for each scenario.[6] Participants could string together these statements using a list of conjunctions and the system would automatically display their proposed support message. The system also allowed participants to edit their support by removing previously added statements and conjunctions. When participants were happy with their support, they progressed to the next scenario. A screen capture of the system is shown in Fig. 5.

Hypothesis: We hypothesise that the categories of emotional support used will differ depending on the stressor.

4.1.2. Results

Categories of emotional support used per stressor: We investigated the categories of emotional support which participants used for each stressor. Fig. 6 shows the results if multiples of the same emotional support category were only counted once (e.g., if a participants gave two PRAISE statements, this was only counted as one). People provided different categories of emotional support depending on the cause of the stress. For example, EMOTIONAL REFLECTION which had been applied in previous related work for stress caused by frustration, as shown in Table 1, was indeed the most commonly used emotional support category in our study for frustration, but used less for say mental demand.

Fig. 7 shows the results if multiples of the same emotional support category were taken into account.[7] This took into consideration that participants were allowed to provide up to four of the same emotional support category in the feedback that they provided. The most frequently used emotional support categories for all stressors have not changed but rather have only been accentuated. For example for the scenario describing Temporal Demand, the emotional support category DIRECTED ACTION (DA) has extended its advantage over the other categories.

To investigate how the use of different support categories compared between different stressors we performed Kruskal–Wallis tests, with stressor scenario as the independent variable, and a numerical dependent variable for each emotional support category, counting how often a participant used that category for that stressor. Multiples that were applied by participants of the same emotional support category were counted in this analysis. This analysis revealed that there were statistically significant differences in the categories of emotional support which people provided based on the stressor that was currently being experienced by the recipient. A statistically significant effect was shown for the following support categories between the different stressors described in our scenarios: DIRECTED ACTION ($\chi(6) = 233.36$; $p < 0.005$), EMOTIONAL ADVICE ($\chi(6) = 50.51$; $p < 0.005$), EMOTIONAL REFLECTION ($\chi(6) = 104.29$; $p < 0.005$), and REASSURANCE ($\chi(6) = 23.18$; $p < 0.01$). These results have been Bonferroni corrected. A statistically significant result was not found for the emotional support category PRAISE.

Next, we explored how the different categories of emotional support provided by our participants varied for each stressor. For this analysis we identified the most frequently used support category for each scenario (stressor), as shown in Fig. 7, and compared it with the other support categories using Wilcoxon Signed Rank tests[8] (Bonferroni corrected). Multiples of the same emotional support category used by participants were included in this analysis (see Fig. 7). EMOTIONAL REFLECTION was shown to have been provided a significantly larger number of times by participants than any other category for the scenarios for which it was the most frequently used: *Interruption*, *Physical Demand*, and *Frustration* ($p < 0.005$). DIRECTED ACTION was also shown to be significantly higher than all others for the scenarios for which it was the most frequently used: *Temporal* and *Mental Demand* ($p < 0.05$). EMOTIONAL ADVICE was found to be significantly higher than only DIRECTED ACTION for *Emotional Demand* ($p < 0.001$), as was REASSURANCE for *Isolation* ($p < 0.001$). This is expected from the data represented in Fig. 7.

These results indicate that people do alter the category of emotional support that they provide depending on the cause of the stress that is being experienced by the recipient of that support. This confirms our hypothesis and supports the differences in the data presented in Figs. 6 and 7.

Combinations of support: Next, we investigated what were the most frequent combinations of emotional support category that people used for each stressor. As the average number of statements used by participants was two (ranging from 2.03 to 2.24), we performed this analysis to a depth of two. To aid readability we only report the combinations for the most frequently used emotional support category per scenario. We began with analysis of the number of occurrences for when the top emotional support category was given solely by itself. Next, we calculated the number of times the leading emotional support category was given multiple times or in combination with other emotional support categories (in any order). These results are reported in Table 6.

4.2. Algorithm generation

From the results of the study, we developed three algorithms for providing emotional support to people experiencing different stressors. These algorithms are presented in Table 7. The following rationale was used for the generation of these algorithms:

(i) For the first algorithm, we determined which emotional support categories were the *most dominantly used* for each stressor (multiple use of an emotional support category in a participants feedback were only counted as one and the categories were analyzed in isolation (i.e. combinations of different support categories were not analyzed at this stage), see Fig. 6). A threshold was set for each category so that more than 50% of participants had used it in their emotional support

[6] The starting screen had zero statements, but asked them to provide at least one.

[7] Means can be seen by dividing the figure's numbers by 100. Standard deviations ranged between 0.48–0.66 for REASSURANCE, 0.5–0.61 for PRAISE, 0.42–0.71 for EMOTIONAL REFLECTION, 0.39–0.78 for EMOTIONAL ADVICE, and 0.17–0.8 for DIRECTED ACTION.

[8] Similar results are obtained using the Signed test, with as only exception that the difference between DIRECTED ACTION and PRAISE is no longer significant for TEMPORAL DEMAND.

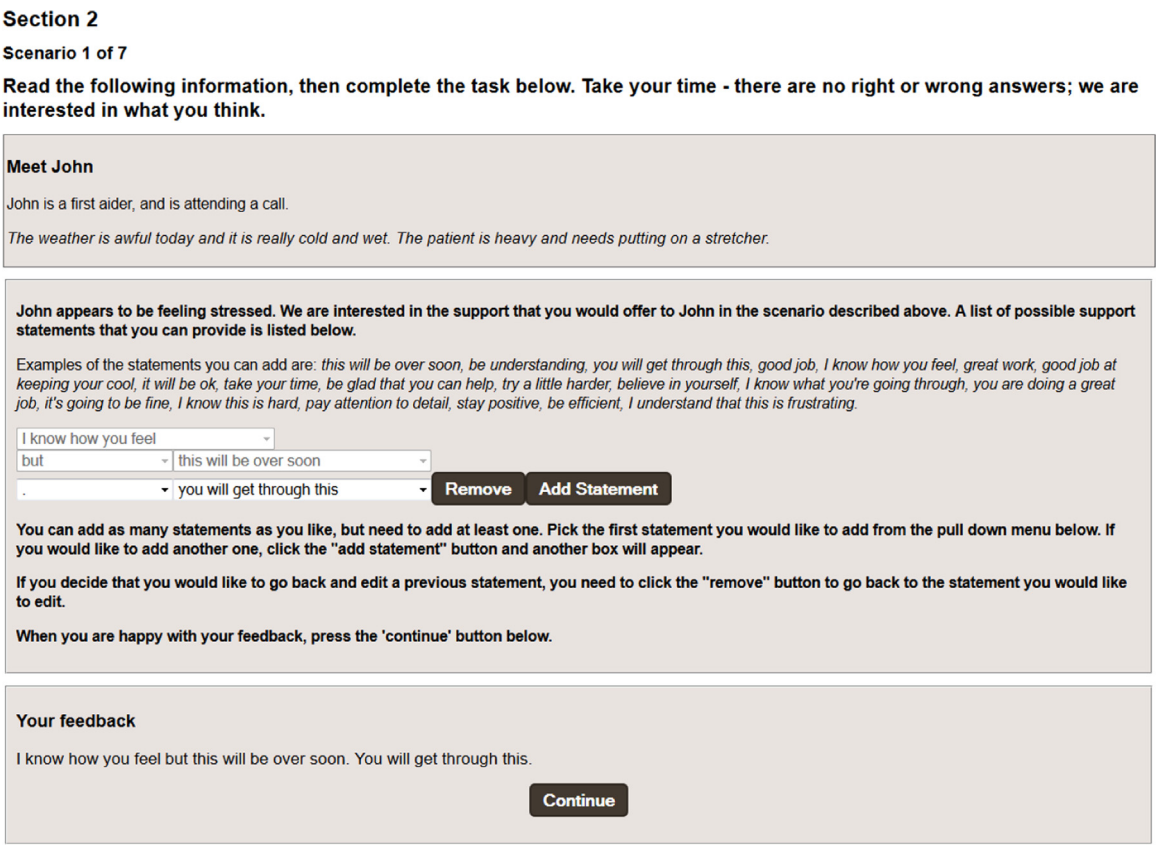

Fig. 5. Screenshot of emotional support algorithm generation study.

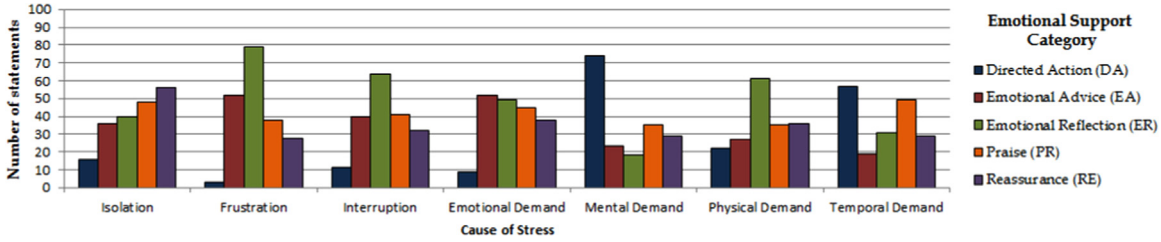

Fig. 6. The number of statements from each support category applied by participants for each scenario. Multiples are not included.

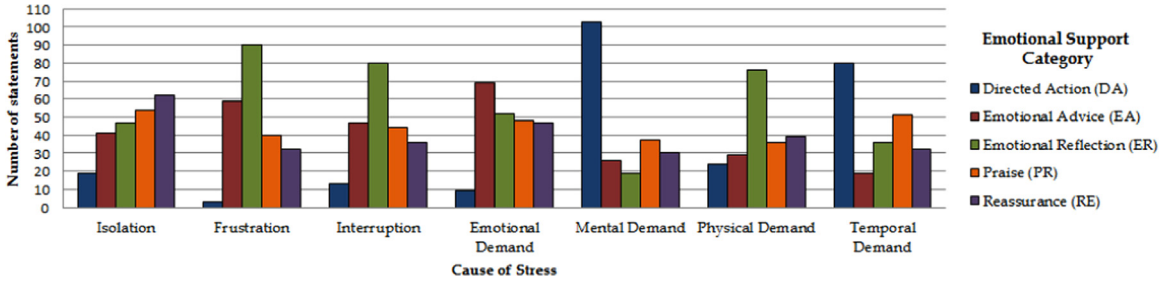

Fig. 7. The number of statements from each support category applied by participants for each scenario. Multiples included.

feedback. A clear example of this was the support category DIRECTED ACTION for *Mental Demand*, as shown in Fig. 6. Algorithm 1 supports five unique emotional support strategies. *Interruption* and *Physical Demand* received the same emotional support, as did *Mental Demand* and *Temporal Demand*.

(ii) For the second algorithm, we took into account the average length of the support feedback provided by participants which consisted of two statements (range of 2.03 to 2.24). With this in mind, our second algorithm had the requirement that the generated emotional support *must contain exactly two emotional support statements*. Algorithm 2 follows the same assumptions as stated for (i), with the addition of *adding a statement from the second most commonly used emotional support category* as applied by participants (see Fig. 6) if the generated emotional support consisted of less than two statements. Algorithm 2 supports four unique emotional support strategies. The same emotional support categories were applied to *Frustration*, *Interruption* and *Emotional Demand*, as well as for *Mental Demand* and *Temporal Demand*.

(iii) For the third algorithm, we explored the combinations of

Table 6
The most frequent combinations of emotional support provided per stressor when used in conjunction with the most frequent emotional support category for that cause (to a depth of 2 to aid readability).

Stressor	Most commonly used emotional support combinations					
Isolation	RE & ER	RE & EA	RE & PR	RE & DA	RE (ONLY)	RE & RE
	22	18	16	10	8	5
Frustration	ER & EA	ER & PR	ER & RE	ER (ONLY)	ER & ER	RE & DA
	40	28	23	10	10	1
Interruption	ER & PR	ER & EA	ER & RE	ER (ONLY)	ER & ER	ER & DA
	23	22	20	15	15	3
Emotional	EA & ER	EA & RE	EA & RE	EA & EA	EA (ONLY)	EA & DA
Demand	21	20	17	15	10	5
Mental	DA (ONLY)	DA & DA	DA & PR	DA & RE	DA & EA	DA & ER
Demand	29	29	22	20	13	9
Physical	ER & PR	ER & RE	ER (ONLY)	ER & ER	ER & EA	ER & DA
Demand	20	19	17	15	14	11
Temporal	DA & DA	DA & PR	DA & RE	DA (ONLY)	DA & ER	RE & EA
Demand	22	21	16	15	13	9

support categories provided by participants. This algorithm followed the same assumptions as stated for (i) and (ii), but instead of using the second most commonly used emotional support category as in (ii), the algorithm uses the *most commonly paired support category* with the most frequently applied category. An exception was made for the rule applied from (ii) for the emotional support category *Mental Demand* so that where one message of an emotional support category was applied the same number of times as multiples of that same category, our algorithm applied the shorter version. In the case of *Mental Demand* this meant that DA was applied instead of DA + DA. The results of the analysis of which were the most frequently used combinations of emotional support was shown in Table 6. Algorithm 3 supports five unique emotional support statements. *Interruption* and *Physical Demand* received the same emotional support, as did *Frustration* and *Emotional Demand*.

4.2.1. Order of support statements chosen by our algorithms

To determine the order in which our algorithms should present the emotional support statements, we analysed the order that the chosen statements had been given by our participants. As a result, a support category that was more commonly applied before another, when given in conjunction with that category, was also provided first by our algorithms. For example, for *Frustration* the most commonly provided categories were EMOTIONAL REFLECTION and EMOTIONAL ADVICE. Participants provided EMOTIONAL REFLECTION before EMOTIONAL ADVICE 38 times in their support, compared to only 3 for the vice versa. The results of this analysis have been applied to our algorithms, as depicted in Table 7.

Table 7
Emotional support algorithms (with Dice Coefficient scores, see formula below).

Stressor	Combinations of emotional support (Dice score)		
	Algorithm 1 >50% once	Algorithm 2 Top 2 statements	Algorithm 3 Highest combinations
Isolation	RE (0.34)	PR + RE (0.48)	RE + ER (0.43)
Frustration	ER + EA (0.8)	ER + EA (0.61)	ER + EA (0.61)
Interruption	ER (0.40)	ER + EA (0.48)	ER + PR (0.50)
Emotional Demand	EA (0.30)	ER + EA (0.46)	ER + EA (0.46)
Mental Demand	DA (0.49)	PR + DA (0.53)	DA (0.49)
Physical Demand	ER (0.39)	ER + RE (0.50)	ER + PR (0.48)
Temporal Demand	DA (0.47)	PR + DA (0.50)	DA + DA (0.53)
Dice Overall	**0.46**	**0.5**	**0.5**

4.2.2. Algorithm accuracy and validation

To investigate how well our proposed algorithms describe the collected data we calculated the Dice coefficient (Deemter et al., 2012) for each algorithm.[9] The Dice coefficient is a well-accepted distance metric for computing the degree of similarity between two sets. We use it here to determine how similar the emotional support generated by our algorithms is to the emotional support provided by the participants. The Dice coefficient was computed by multiplying the number of emotional support categories that the two support messages had in common by two, divided by the overall number of statements in each:

$$dice(D_H, D_A) = \frac{2 \times |D_H \bigcap D_A|}{|D_H| + |D_A|},$$

where D_A is the set of emotional support categories that our algorithm used and D_H the set of emotional support categories that the participant selected. As the formula uses sets, duplicates of the same support category are ignored (i.e. a message with two praise statements and a message with one praise statement are treated the same). For example, if our algorithm picked a DIRECTED ACTION and a PRAISE statement and a participant chose to offer DIRECTED ACTION, PRAISE as well as EMOTIONAL REASSURANCE, the Dice coefficient score would be: 2*2/(2+3)=0.8. Dice coefficient scores range from 0 to 1, where 1 indicates a perfect match with the data. The Dice coefficient scores for our algorithms are shown in Table 7. From the scores, we can see that Algorithm 2 and Algorithm 3 match the data gathered in this study slightly better than Algorithm 1. Although the scores for these algorithms were promising, we could not determine from this data which was the more effective algorithm for generating emotional support.

4.2.3. Selecting individual statements

The algorithms so far only prescribe which emotional support categories to use and in which order. To apply the algorithms, and be able to evaluate their effectiveness in future studies, a selection also needs to be made of which statement to use within a category. It is the intention that a virtual agent in future would be able to choose a random statement from a set of statements for each category. However, while we have the statements produced by the crowd-sourcing, we do not know whether they would all be as effective. We decided to use participants' frequency of use as an estimation of effectiveness, assuming that statements that had been used more frequently would be more effective.

To determine whether certain statements were better suited to particular stressors than others within the categories, we analysed which statements within each category were more commonly used. Table 8 shows that certain statements were far more frequently used than others. For example, for Mental Demand, "*Try a little harder*" was never used while two other DIRECTED ACTION statements were used very frequently. We decided that the algorithms would use the most frequently used statement in the category for that stressor (as indicated in Table 8). For example, for Isolation, the reassurance used by the algorithms will be "*It's going to be fine.*". This choice was made to ensure the comparison between algorithms would be fair, with each algorithm using statements that had been frequently used by participants for each stressor.

[9] We are only measuring here how well the algorithms describe the collected data rather than measuring how well they would predict new data. Alternatively, we could have split the data into two parts, based the algorithms on one part and measured the Dice score on the other part. However, this would have required more participants.

Table 8
Frequency of emotional support statements used. Categories not used by any of the algorithms for a scenario greyed out. Most frequently used statement(s) in a used category for a scenario in bold. Italicised statement was chosen when two were used frequently.

Category	Statement	Frequency used per scenario						
		PD	TD	MD	IN	FR	ED	IS
DA	Pay attention to detail	2	23	**52**	6	0	1	7
	Take your time	16	18	47	5	3	6	11
	Be efficient	1	**38**	4	2	0	2	1
	Try a little harder	5	1	0	0	0	0	0
EA	Stay positive	18	9	6	13	16	**26**	22
	Be glad that you can help	4	1	2	6	**38**	11	7
	Believe in yourself	2	9	16	2	3	12	10
	Be understanding	5	0	1	**25**	1	20	2
ER	I know how you feel	4	0	2	5	8	7	4
	I understand that this is frustrating	17	11	6	**32**	**68**	5	13
	I know what you are going through	4	3	2	5	5	6	0
	I know this is hard	**29**	11	8	17	6	**30**	**15**
PR	Great work	2	4	6	0	1	0	6
	Good job	7	11	9	**31**	7	18	14
	You are doing a great job	**27**	**36**	**22**	13	32	28	**34**
	Good job at keeping your cool	6	9	8	***31***	5	16	13
RE	It is going to be fine	10	8	9	6	10	7	**30**
	It will be OK	8	6	4	9	8	13	13
	You will get through this	***21***	18	17	21	13	27	18
	This will be over soon	**21**	10	1	21	3	4	15

To determine which conjunctions to use to connect the statements selected by our algorithms, we investigated the most commonly used conjunctions when participants had used the same combination of categories. These were used to produce the combined statements in Table 9.

4.3. Study 5: evaluation and refinement of algorithms

In this study we evaluated the emotional support generated by the algorithms for the stressors described in the scenarios. This was to determine which algorithm generated the most appropriate emotional support for each individual stressor. The most appropriate strategy for each of the stressors would then be applied to create a new and improved algorithm. To achieve this, a different approach to how participants evaluated the generated emotional support statements was taken. Participants were asked to imagine that they were the person described in the presented scenario and to rate the generated emotional support messages presented to them. This signifies an important change as to how participants evaluated the statements, as the emotional support that one person may give to a stressed individual may differ from the emotional support that they would wish to receive if they were that stressed individual.

4.3.1. Method

Participants: As before, we used Amazon's Mechanical Turk and the same methodology and necessary requirements for the recruitment of participants. One hundred participants completed the study. 51% were female and 49% male. 20% were aged between 18 and 25, 49% aged between 26 and 40 and 31% aged between 41 and 65. The average completion time was 6.5 min. Participants were paid $1.

Procedure: Participants were shown the seven previously validated scenarios (in randomised order), but with a slight variation. For this study, instead of describing another person in the scenario (i.e. 'Mark is a first aider'), participants were told to imagine themselves in the scenario in place of the first aider. For each scenario, participants were shown the unique statements generated by the algorithms for that scenario (as shown in Table 9) in randomized order. For example, if an 'algorithm A' and an 'algorithm B' produced the same emotional support for the same stressor then only one message would be shown for participants to rate. They rated each statement on four scales measuring APPROPRIATENESS, EFFECTIVENESS, HELPFULNESS, and SENSITIVITY. These scales have been previously applied to validate emotional support cf. (Jones and Burleson, 1997). Responses for each scale were recorded on a Likert scale ranging from 1 to 9. A screen capture is shown in Fig. 8.

Table 9
Mean ratings for emotional support messages generated by the algorithms. Shaded *grey* indicates the highest rated message for that stressor and which algorithms generated that combination of Emotional Support.

Stress caused by	Message	Comprised of	App	Eff	Helpf	Sens	Alg
Isolation	It is going to be fine[a]	RE	6.12*	5.01	4.98	5.96*	1
	You are doing a great job, it is going to be fine[a]	PR + RE	6.97*	6.48*	6.43*	6.84*	2
	It is going to be fine and I know this is hard[a]	RE + ER	6.58*	5.91*	5.90*	6.40*	3
Frustration	I understand that this is frustrating but be glad that you can help[a]	ER + EA	6.50*	5.63	5.69*	6.06*	1,2,3
Interruption	I understand that this is frustrating[a]	ER	6.21*	5.05	4.94	6.07*	1
	I understand that this is frustrating however be understanding	ER + EA	5.75*	4.95	4.89	5.25	2
	I understand that this is frustrating however good job at keeping your cool[a]	ER + PR	7.23*	6.97*	6.87*	7.09*	3
Emotional Demand	Stay positive	EA	6.13*	4.93	4.92	5.39	1
	I know this is hard but stay positive[a]	ER + EA	6.52*	5.88*	5.81*	6.29*	2,3
Mental Demand	Pay attention to detail	DA	5.68*	4.82	4.65	4.26	1,3
	You are doing a great job but pay attention to detail[a]	PR + DA	6.19*	5.77*	5.64	5.69*	2
Physical Demand	I know this is hard	ER	6.02*	4.74	4.48	5.59*	1
	I know this is hard but you will get through this[a]	ER + RE	6.85*	6.23*	6.07*	6.61*	2
	I know this is hard but you are doing a great job[a]	ER + PR	7.44*	6.89*	6.77*	7.29*	3
Temporal Demand	Be efficient	DA	4.77	3.83	3.60	3.53	1
	You are doing a great job but be efficient	PR + DA	5.64*	4.94	4.84	5.02	2
	Be efficient but pay attention to detail	DA + DA	5.43	4.71	4.62	4.42	3

App = Appropriateness, Eff = Effectiveness, Helpf = Helpfulness, Sens = Sensitivity, Alg = Algorithms.

[a] For a message indicates significance when all four scales for that strategy were counted together, results are Bonferroni corrected.

* Denotes a significance of $p < 0.001$ for Z-scores from a mean of 5.

4.3.2. Results

The independent variables in the analysis are scenario and emotional support message (as selected by the algorithms). The results of the ratings which participants gave for each message are presented in Table 9. There was a clear winning (received the highest score) support message for each scenario, as reported on each of the four scales. The highest rated message for each scenario is presented in Table 10.

For the scenario describing *Physical Demand*, a MANOVA showed there was a significant effect of emotional support message on the four dependent variables (our scales), Pillai's trace $F(8, 590)=7.58$, $p < 0.01$, partial $\eta^2=0.09$.

Each dependent variable was subjected to a further ANOVA analysis in order to show whether this trend was the same for each of the separate dependent variables. For the measure of difference between the scale APPROPRIATENESS and the other three scales, an ANOVA showed there was an overall significant difference between the means, $F(2, 297)=18.18$, $p < 0.01$, partial $\eta^2=0.1$. Similarly, significant differences were found for: EFFECTIVENESS ($F(2, 297)=29.31$, $p < 0.01$, partial $\eta^2=0.16$), HELPFULNESS ($F(2, 297)=30.5$, $p < 0.01$, partial $\eta^2=0.17$) and SENSITIVITY ($F(2, 297)=22.88$, $p < 0.01$, partial $\eta^2=0.13$).

Similar conclusions could be drawn for the following stressors which were described in our scenarios:

Temporal Demand: Pillai's trace $F(8, 590)=4.3$, $p < 0.01$, partial $\eta^2=0.06$. APPROPRIATENESS ($F(2, 297)=4.9$, $p < 0.01$, partial $\eta^2=0.03$), EFFECTIVENESS ($F(2, 297)=7.63$, $p < 0.01$, partial $\eta^2=0.05$), HELPFULNESS ($F(2, 297)=9.37$, $p < 0.01$, partial $\eta^2=0.06$), SENSITIVITY ($F(2, 297)=14.68$, $p < 0.01$, partial $\eta^2=0.09$).

Interruption: Pillai's trace $F(8, 590)=9.4$, $p < 0.01$, partial $\eta^2=0.11$. APPROPRIATENESS ($F(2, 297)=4.9$, $p < 0.01$, partial $\eta^2=0.03$), EFFECTIVENESS ($F(2, 297)=7.63$, $p < 0.01$, partial $\eta^2=0.05$), HELPFULNESS ($F(2, 297)=9.37$, $p < 0.01$, partial $\eta^2=0.06$), SENSITIVITY ($F(2, 297)=14.68$, $p < 0.01$, partial $\eta^2=0.09$).

Mental Demand: Pillai's trace $F(4, 193)=10.24$, $p < 0.01$, partial $\eta^2=0.18$. APPROPRIATENESS ($F(1, 196)=2.97$, $p < 0.01$, partial $\eta^2=0.02$), EFFECTIVENESS ($F(1, 196)=9.01$, $p < 0.01$, partial $\eta^2=0.04$), HELPFULNESS ($F(1, 196)=9.36$, $p < 0.01$, partial $\eta^2=0.05$), SENSITIVITY ($F(1, 196)=26.37$, $p < 0.01$, partial $\eta^2=0.11$).

Despite a significant effect on emotional support message being reported for both Emotional Demand and Isolation ($p < 0.005$), no significant comparisons were reported as to whether the trends for each scale were similar. No comparisons could be made for the stressor Frustration as it considered only one unique emotional support message.

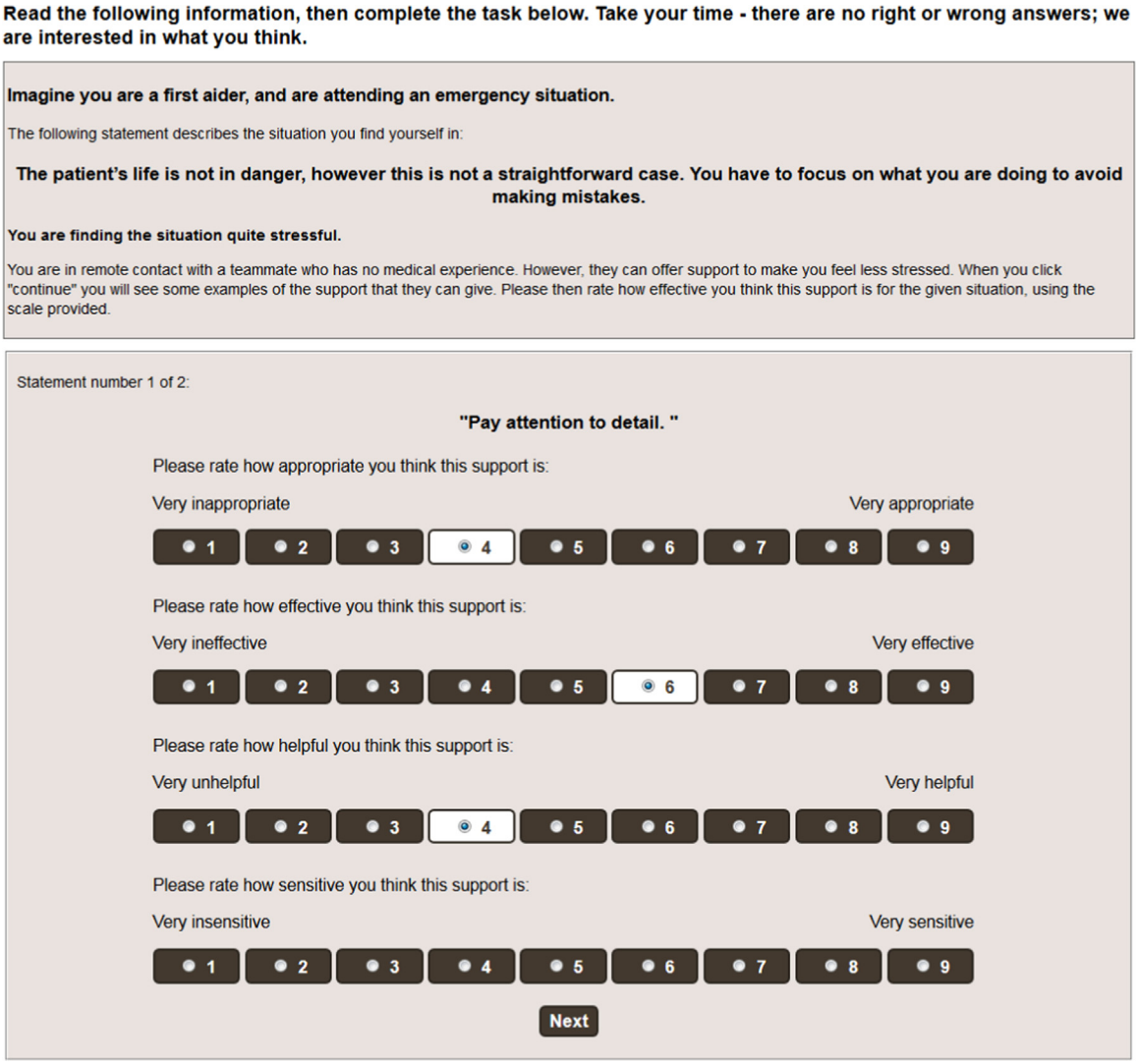

Fig. 8. Screen capture of system used in the 'Refinement of Algorithms' study.

Table 10
Refined emotional support algorithm. Support categories applied for each scenario and corresponding top emotional support messages.

Scenario (stressor)	Categories	Message
Temporal Demand	PR + DA	You are doing a great job but be efficient
Physical Demand	ER + PR	I know this is hard but you are doing a great job
Mental Demand	PR + DA	You are doing a great job but pay attention to detail
Emotional Demand	ER + EA	I know this is hard but stay positive
Interruption	ER + PR	I understand that this is frustrating however good job at keeping your cool
Frustration	ER + EA	I understand that this is frustrating but be glad that you can help
Isolation	PR + RE	You are doing a great job, it is going to be fine

The results indicate that Algorithm 2 was marginally better at generating emotional support than Algorithm 3, with five of the highest rated messages being generated by Algorithm 2 compared to four being generated by Algorithm 3.[10] This is shown in Table 9. Algorithm 1 was shown to perform relatively poorly with only one of the highest rated emotional support messages being generated by it. The emotional support generated in that instance was also generated by Algorithm 2 and 3, therefore rendering Algorithm 1 unnecessary.

4.3.2.1. Scale correlation. Interestingly, the winning support strategy for each scenario was ranked as having the highest mean values on all four scales. This posed the question as to whether participants perceived there to be any distinction between the measurements of each individual scale. Using Cronbach's Alpha analysis the alpha reliability of the four scales was 0.94, indicating the scales to have a very strong reliability. An alpha of .70 or above is considered satisfactory. This indicated that when a participant rated a statement highly (or lowly) on one scale, they also rated it similarly on the other scales. The scale measuring 'EFFECTIVENESS' was the most strongly correlated with the others. As the alpha value is very high, it also suggests that there is a high level of item redundancy; that is, participants are not making a clear distinction between the scale measures and are therefore rating them similarly. This would suggest that not all of the four scales are regarded in further studies, and the scale could be reduced to three items. These results are presented in Table 11.

4.3.3. Optimized algorithm

Based on the study results, a refined algorithm has been designed which produces emotional support considered the best of those produced by our previous algorithms. The emotional support categories which the '*optimized*' algorithm applies, and the messages used to instantiate these categories (using the same rationale as above for using the most frequently used messages) are shown in Table 10. The results of the study (as presented in Table 9) showed that the messages applied by our optimized algorithm were considered to perform well (above average on the scales for Appropriateness, Effectiveness, Helpfulness and Sensitivity).

The refined algorithm uses four unique emotional support strategies as being the most suitable for our identified stressors (instantiated using seven unique emotional support messages). For Isolation, it uses support comprised of *Praise* and *Reassurance*. For Physical Demand and Interruption, it uses support comprised of *Emotional Reflection* and *Praise*. For Temporal Demand and Mental Demand, it uses support comprised of *Praise* and *Directed Action*. For Frustration and Emotional Demand, it uses support comprised of *Emotional Reflection* and *Emotional Advice*.

Three out of the four unique emotional support strategies include the emotional support category *Praise*. This suggests that *Praise* is a suitable form of emotional support for most stressors. Between all of the emotional support strategies, each one of our defined emotional support categories was used.

[10] Due to the lack of significance reported in the Temporal Demand scenario, it could be argued that both Algorithm 2 and 3 produce the same number of highest rated messages. However, as Algorithm 2 receives the best scores for each scale for Temporal Demand, we consider this algorithm to be the marginally better one.

Table 11
Inter-item correlation matrix.

	Appropriateness	Effectiveness	Helpfulness	Sensitivity
Appropriateness	1	0.766	0.745	0.792
Effectiveness	0.766	1	0.937	0.757
Helpfulness	0.745	0.937	1	0.767
Sensitivity	0.792	0.757	0.767	1

4.4. Study 6: evaluation of the final emotional support algorithm

The previous studies have led to the development of an emotional support algorithm that has been evaluated to provide the most suitable emotional support messages from a corpus of statements which a human might provide for individual stressors. Following the development of our algorithm, this study aimed to investigate whether the generated emotional support by our algorithm was uniquely suitable to individual stressors (depicted by our scenarios) or if there are more dominant and general emotional support paradigms which people choose to apply. In other words, if the emotional support generated by our algorithm was solely appropriate for the stressor that it was specifically tailored for or whether it was considered to be appropriate for other stressors.

This study presented participants with multiple scenarios each validated to describe a unique main stressor. Participants were then presented with each of the unique emotional support messages generated by our algorithm and were asked to rate how effective they considered the message to be for that each scenario. This study acted as an evaluation of our algorithm and investigated whether there are genuine differences between the emotional support strategies used by our algorithm when given for different stressors. This study sought to provide evidence as to how well our emotional support algorithm performs.

4.4.1. Method

Participants: As before, we used Amazon's Mechanical Turk and the same methodology and necessary requirements for the recruitment of participants. One hundred participants completed the study. 54% were male and 46% female. 13% were aged between 18 and 25, 55% aged between 26 and 40, 30% aged between 41 and 65, and 2% aged over 65. The average completion time was 7.5 min. Participants were paid $0.70.

Procedure: Participants were shown the seven scenarios, each having been previously validated to depict a unique stressor, in randomised order. They were presented with the seven emotional support messages generated by our refined algorithm (see Table 10) in randomised order and rated how effective they considered each message to be for the current scenario.[11] To reduce

[11] An alternative would have been to ask participants to choose the best message per scenario, but that would have provided less insight into participants' opinions (for example, not showing when two messages were regarded as almost as good), and may have reduced the effort spent on considering individual messages.

the time needed to complete the study (given participants needed to provide 49 judgements), we have only applied the most strongly correlated scale[12] from the four scale measures we used previously for the measurement of participants' ratings during this study.[13] This was the scale which measured 'EFFECTIVENESS'. As in the previous studies, this scale ranged from 1 (*very ineffective*) to 9 (*very effective*). Participants were again asked to imagine that they were the person depicted in the scenario and were the receiver of the support rather than the provider. The seven emotional support messages were presented one after the other for each scenario, in a randomised order. A screen shot is shown in Fig. 9.

4.4.2. Results

The median effectiveness rating for each message for each scenario is presented in Fig. 10. The *star* presented in Fig. 10 indicates which emotional support message was intended for that stressor. From the data presented in this chart, it is clear that the message originally considered to be the most suitable for *Physical Demand* has been universally considered as being the most (or equivalently as) effective across all scenarios by our participants. This message consisted of categories *Emotional Reflection* and *Praise* and is shown in Table 10. However, as can be observed from the chart in Fig. 10, other emotional support messages have been found to be as equally effective for their intended stressors. This is true for the emotional support strategies designed for *Interruption* and *Isolation*, with the strategy designed for *Physical Demand* also having been considered to be the most effective for its intended stressor.

Between individual emotional support messages: There was a clear distinction between emotional support messages in terms of EFFECTIVENESS for all of the scenarios. A one-way unrelated analysis of variance was performed for each scenario. An overall significant effect between the rating of messages was found for all scenarios: Physical Demand ($F(6, 693)=42.23$, $p < 0.001$), Temporal Demand ($F(6, 693)=24.40$, $p < 0.001$), Mental Demand ($F(6, 693)=13.86$, $p < 0.001$), Emotional Demand ($F(6, 693)=31.50$, $p < 0.001$), Interruption ($F(6, 693)=39.37$, $p < 0.001$), Frustration ($F(6, 693)=36.40$, $p < 0.001$), and Isolation ($F(6, 693)=53.58$, $p < 0.001$). Table 12 presents the results from Scheffé's test[14] for each scenario showing the significant differences between the generated emotional support messages in each scenario.

Emotional support message discussion: Out of the seven emotional support messages provided by our algorithm, three were rated highest for their corresponding scenario (the scenario for which they were designed), namely the messages intended for Physical Demand, Interruption and Isolation. The three messages intended for Emotional Demand, Mental Demand and Frustration performed sufficiently well; they were not found to be significantly worse than other messages in their corresponding scenario. Only the message intended for Temporal Demand performed badly; significantly worse than several other messages in the Temporal Demand scenario.

Interestingly, the message intended for when the recipient experienced Physical Demand was rated consistently highly across all scenarios. This would suggest that the emotional support message generation by our algorithm for Physical Demand is applicable for all of our described stressful scenarios.

Despite not having been rated as the most effective emotional support message for its intended stressor (though not significantly worse than any other), the message designed for Mental Demand was still rated as more effective for that stressor than any other one. The reason for it not being rated higher may have been due to the wording of the message rather than the emotional support categories that it was comprised of.

Similar results were obtained for the message intended for Frustration, with the message's highest rating coming in the scenario describing Frustration as well as in the scenario describing Isolation. These results may indicate a problem with the Isolation scenario describing not solely isolation but also frustration (this scenario only validated with marginal kappa).

Interestingly, the emotional support message intended for Emotional Demand had a similarly high effectiveness across all scenarios. This may indicate that emotional support intended for Emotional Demand is universally appropriate and effective for all types of stress (if not the most effective).

No interesting trends were observed for the emotional support message intended for Temporal Demand, with it being rated consistently lowly across scenarios.

Rather than just focusing on the individual emotional support messages (i.e. the wording used) generated by our algorithm we also investigate the differences found between strategies (i.e. the emotional support categories used to obtain the statements for the message) comprised of the same categories. Despite our messages all being generated with validated emotional support statements, there may still be statement bias for certain stressors, where one keyword may be considered more suitable than another. With this in mind, we then looked to explore how the different emotional support strategies were rated by participants.

Between emotional support strategies: While the results above indicate that the algorithm is performing well (except for Temporal Demand), this result is linked to the messages used to instantiate the support categories, and other messages may have had worse results. To investigate this further, we also considered the results per emotional support strategy. As discussed above, the algorithm generated messages using four unique emotional support strategies, as listed below and indicated by colour in Fig. 10.

Strategy 1 (PR+RE): Comprises of emotional support from the categories PRAISE and REASSURANCE. This strategy is used for *Isolation* and represented by the colour *Green*.

Strategy 2 (ER+EA): Comprises of emotional support from the categories EMOTIONAL REFLECTION and EMOTIONAL ADVICE. This strategy is used for *Frustration* and *Emotional Demand* and represented by the colour *Red*.

Strategy 3 (PR+DA): Comprises of emotional support from the categories PRAISE and DIRECTED ACTION. This strategy is used for *Mental Demand* and *Temporal Demand* and represented by the colour *Orange*.

Strategy 4 (ER+PR): Comprises of emotional support from the categories EMOTIONAL REFLECTION and PRAISE. This strategy is used for *Interruption* and *Physical Demand* and represented by the colour *Blue*.

The results of a statistical analysis (ANOVA) for the difference of user rated *effectiveness* for these strategies are reported in Table 13. To aid readability, only statistically significant results are shown.

Emotional support strategy discussion: The comparison between the different emotional support strategies used by our algorithm for different stressors provided some interesting conclusions.

Firstly, reflecting on the effectiveness of the algorithm, ER+PR performed well for both stressors it had been designed for, namely Physical Demand and Interruption, independent of the message

[12] In the sense of having the highest average correlation with the three other scales.

[13] The four measures were strongly correlated, with a very high Cronbach's Alpha value of 0.94. While we would have preferred to have kept three measures, the need to reduce the study time to one we were comfortable using in Mechanical Turk meant that we decided to only use one measure.

[14] Scheffé's method is a single-step multiple comparison procedure for judging all contrasts in an analysis of variance (Scheffe, 1953).

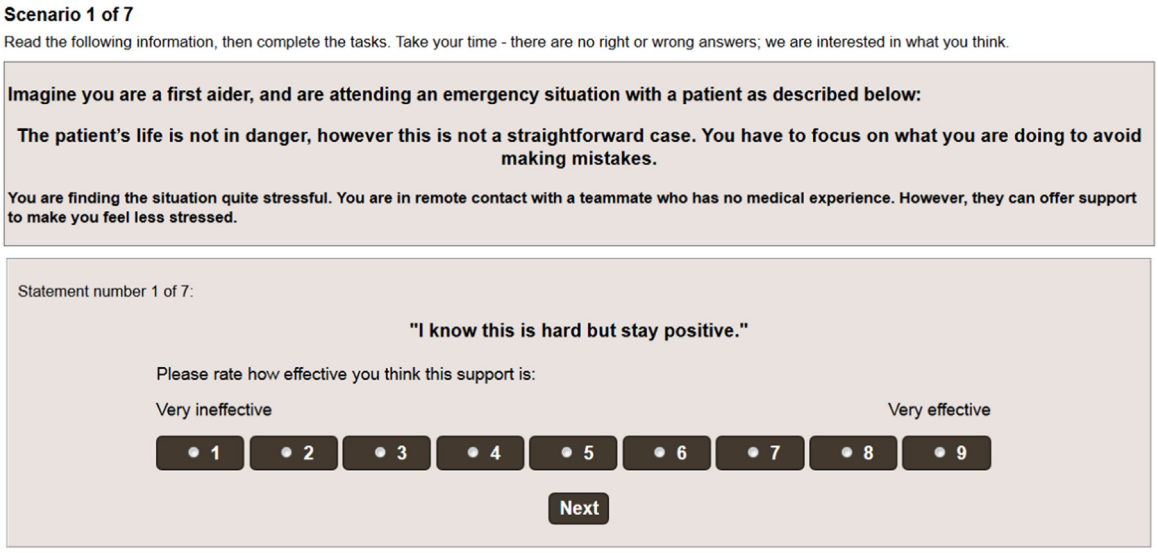

Fig. 9. Screen capture of system used in the 'Evaluation of a Final Emotional Support Algorithm' study.

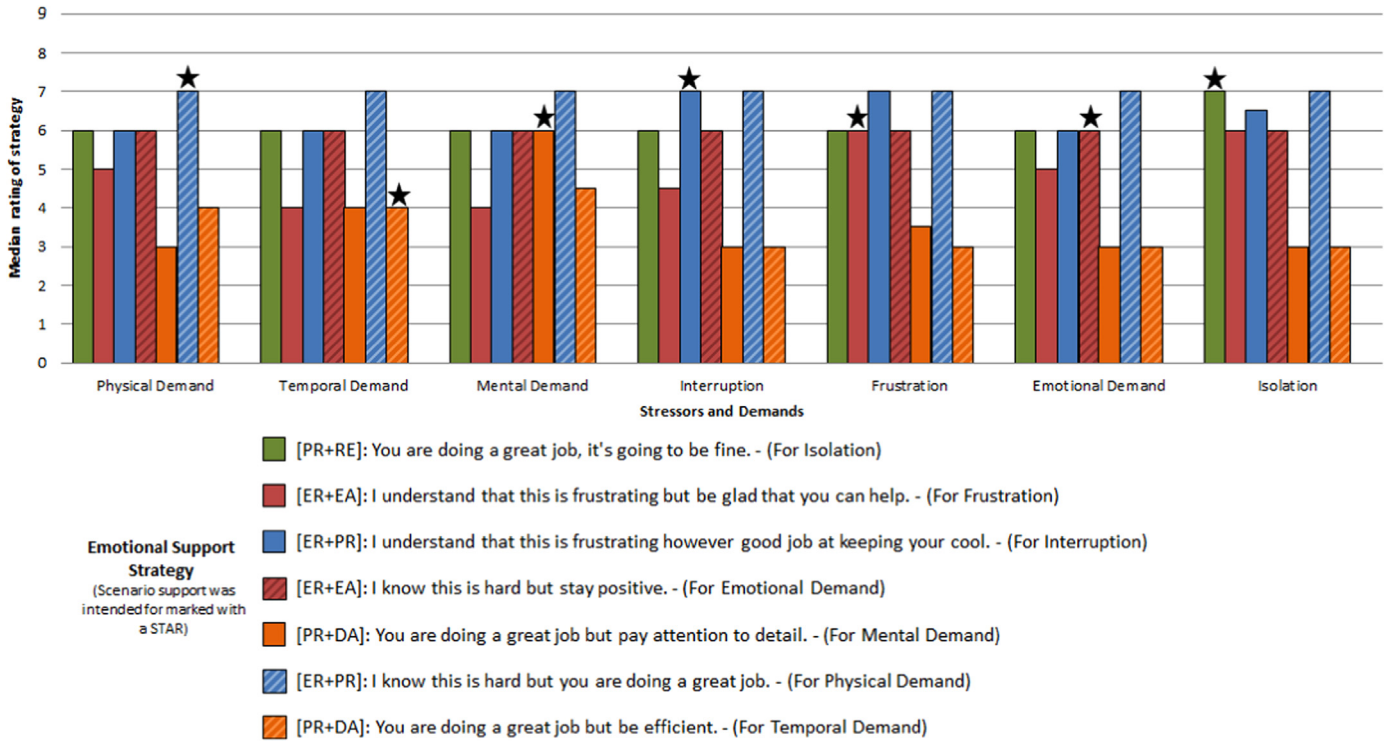

Fig. 10. Emotional support messages rated per scenario. Messages represented by the same colour indicates that message comprised of the same emotional support categories. Star indicates message designed for that scenario.

used. It performed better for those stressors than ER+EA and PR+DA, and as well as PR+RE. ER+EA performed well for both stressors it had been designed for, namely Frustration and Emotional Demand. It performed better than PR+DA, as good as PR+RE, as good as ER+PR on Frustration, but not as good on Emotional Demand where one of its messages performed better than the other. Only one message was used for PR+RE, so we cannot reflect on the influence of its message wording. PR+DA performed similarly badly on all stressors with the exception of Mental Demand independent of the message used. However, for Mental Demand, the PR+DA message designed for Temporal Demand was clearly less effective than the message designed for Mental Demand. So, overall, we see some influence of message choice, but also evidence that different messages within a category can perform very similarly.

The significant difference between PR+DA and all other strategies for all stressors except for Mental Demand for ER+EA, shows that PR+DA was considered relatively ineffective for all stressors expect for Mental Demand (one of the two stressors for which it was designed, the other being Temporal Demand). PR+DA was the only strategy including *directed action* suggesting that although people consider it appropriate to give *directed action* to those experiencing stress (as validated through our algorithm generation process), they may not necessarily wish to receive *directed action* when experiencing stress. This hypothesis has some support from the literature as studies have suggested that the emotional support that people provide to a stressed individual may sometimes be considered

Table 12
Results of Scheffé's test denoting the significance in differences between Emotional Support Messages for each Scenario. FR=Frustration, PD=Physical Demand, IS=Isolation, INT=Interruption, ED=Emotional Demand, MD=Mental Demand, TD=Temporal Demand.

Scenario	Message intended for	FR	INT	ED	MD	PD	TD
Physical Demand	IS	*			***		***
	FR		*		***	***	***
	INT				***		***
	ED				***	**	
	MD					***	
	PD						***
Temporal Demand	IS	***			***		***
	FR		***	*		***	
	INT				***		***
	ED				*		*
	MD					***	
	PD						***
Mental Demand	IS	***					***
	FR		*		*	***	
	INT						**
	ED					*	
	MD						*
	PD						***
Emotional Demand	IS				***		***
	FR		*	*		***	***
	INT				***		***
	ED				***		***
	MD					***	
	PD						***
Interruption	IS	*			***		***
	FR		***	*	*	***	*
	INT				***		***
	ED				***		***
	MD					***	
	PD						***
Frustration	IS				***		***
	FR				***		***
	INT				***		***
	ED				***		***
	MD					***	
	PD						***
Isolation	IS	***			***		***
	FR		*		***	***	***
	INT				***		***
	ED				***		***
	MD					***	
	PD						***

* $p < 0.05$.
** $p < 0.005$.
*** $p < 0.001$.

insensitive, unhelpful or inappropriate by the receiver (Lehman and Hemphill, 1990). The exception found for Mental Demand may indicate that people only wish to receive *directed action* when the stress they are experiencing is down to their perceived abilities to cope. Although the strategy was rated as performing relatively well for this stressor, the other strategies were as well, suggesting that those are equally appropriate. Therefore, we can conclude that participants have rated this emotional support strategy on the whole as ineffective.

The lack of significant differences between PR+RE and ER+PR for any stressor may indicate that people find *reassurance* and *emotional reflection* equally effective in stressful situations. In contrast, the significant differences for every stressor between PR+RE and ER+PR on the one hand and PR+DA on the other hand shows people perceived the effectiveness of *directed action* as lower compared to *reassurance* and *emotional reflection*, when used with *praise*.

Strategy 2 is interesting as it is the only support strategy that does not include *praise*. Based on the results in Fig. 10 and Table 13, ER+EA, which consisted of *emotional reflection* and *reassurance*, was considered to be reasonably effective across all scenarios.

Finally, it needs to be considered that when participants generated emotional support messages, they were presented with many possibilities to choose from. This enabled them to create plausible messages, but we cannot assume that they would always have created the most optimal message. However, when we asked participants to make a decision between a much small number of possibilities, as in this algorithm evaluation study, this became a much easier task for people to achieve reliably.

4.5. Limitations

A limitation with the final study and with research of this nature is that people tend to rate strategies on the exact wording and phrasing of the message, not just on the categories of support that it contains. Despite our attempts to ensure that each strategy was comprised of previously validated statements, categorized into specific emotional support categories, it remains hard to guarantee that the wording or phrasing of one statement when used in conjunction with another might not have unintentionally made the message become applicable to other stressors. Whereas other generated messages using different statements from our emotional support categories may have been inadvertently considered as being better suited as support for the stressor which another message was intended for. This emphasizes the potential limitation of a using a limited set of examples for each emotional support strategy.

Another limitation of this study and the previous ones is that each stressor has only been depicted through one scenario. It is possible that the support strategies participants used, and their ratings of support strategies and messages may have been influenced by the wordings of the specific scenarios. We do not expect that it would influence the choice of strategies (though this warrants investigation), but it may well influence the choice of particular messages within the category. For example, one message, namely the 'be glad that you can help' one used for EA, seems linked to the scenario (where the CFR is helping a patient) and may not be appropriate for other frustration situations. The others used seem quite generic.

Caution should also be taken when drawing conclusions about how appropriate the emotional support generated for Frustration is by our algorithm. This is because the same emotional support strategy has been applied throughout the algorithm refinement process as each of our three original algorithms applied the same strategy. Similarly, when drawing conclusions about how appropriate the emotional support strategies are, it should be noted that the strategy PR+RE only has one instance (as it was only intended for Isolation) and so was not subjected to comparisons. With only one instance the potential for participants not liking the specific use of language employed by the strategy increases.

A clear limitation of all studies described in this section is that they took place in artificial situations. This meant that participants were under no stress when rating the emotional support messages and may have had difficulty imagining the described scenarios. How one thinks one would react to support in a situation may differ from how one actually reacts if that situation were a reality.

Table 13
Differences between emotional support strategies for each individual stressor.

	Strategy 2 ER + EA	**Strategy 3** PR + DA	**Strategy 4** ER + PR
Strategy 1 PR + RE	(Physical Demand)**(Temporal Demand, Mental Demand, Isolation)***	(Mental Demand)**(Physical Demand, Temporal Demand, Emotional Demand, Interruption, Frustration, Isolation)***	
Strategy 2 ER + EA		(Temporal Demand)*(Physical Demand, Emotional Demand, Interruption, Frustration, Isolation)***	(Physical Demand, Temporal Demand, Mental Demand, Emotional Demand, Interruption, Isolation)***
Strategy 3 PR + DA			(Physical Demand, Temporal Demand, Mental Demand, Emotional Demand, Interruption, Frustration, Isolation)***

* $p < 0.05$.
** $p < 0.005$.
*** $p < 0.001$.

4.6. Summary

In this section, we investigated how people provide emotional support for recipients who are experiencing different stressors. This led to the development of an algorithm which could be used by an embodied agent to provide appropriate emotional support depending on the source of the recipient's stress. This process began with the generation of three algorithms, as detailed in Section 4.1. These algorithms were then refined to produce one algorithm through the process of applying what was considered as the best emotional support strategy used by the algorithms for each of our identified stressors. This process was fulfilled by the study detailed in Section 4.3. The performance of this algorithm was then evaluated in the study presented in Section 4.4. The emotional support messages which were produced by our algorithm and the emotional support strategies which it used, were assessed in the study. We assessed whether the emotional support messages and strategies that had been produced for each stressor were unique to that individual stressor, or if some messages and strategies where equally suitable for those which they were not designed for. The results of the study revealed that some of the generated messages were appropriate for more than the stressor that they had been designed for. Equally the results demonstrated that certain emotional support messages and strategies are more appropriate for individual stressors. These results have provided empirical evidence to support the answer to the research question: What are the different strategies for providing emotional support to people experiencing different stressors? The emotional support strategy comprising of *emotional reflection* and *praise* was rated highly not only for its intended stressor for which it was designed (Interruption and Physical Demand) but was also rated highly for all the other identified stressors. Similarly, the emotional support category designed for Isolation (comprised of *praise* and *reassurance*) was rated, not only as performing very well for its intended stressors but as performing consistently well across all the other identified stressors. Although not as comprehensive, the emotional support categories designed for Emotional Demand and Frustration (comprised of *emotional reflection* and *emotional advice*) and for Mental Demand (comprised of *praise* and *directed action*) also performed adequately for the stressors for which they were designed. Importantly the emotional support strategy designed for Temporal Demand was not considered to have performed well, for its intended stressor or for any other stressor (despite being comprised of the same emotional support categories as applied for Mental Demand).

These conclusions signify an important step for the development of a virtual agent (potentially perceived as a teammate) which seeks to provide appropriate and effective emotional support.

5. Conclusions and future work

This paper presents the empirically-led development and the evaluation of an emotional support algorithm that tailors its support to the stressor experienced.

5.1. Main results and lessons learned

Table 14 summarizes the six studies presented and their main results.

In Phase 1 of the research, we first developed and validated seven scenarios, each of which describes a situation that exhibits an individual stressor. These validated scenarios enabled a study into how people provide emotional support to individuals experiencing different stressors, resulting in a corpus of emotional support statements. We identified emotional support categories and validated a subset of statements as strongly belonging to a particular category.

In Phase 1, we learned that crowd-sourcing is a rapid and effective way to (1) rapidly generate a corpus of emotional support statements for scenarios, (2) validate that a scenario depicts a particular stressor for the general public, and (3) validate that an emotional support statement belongs to a particular emotional support category. As we have shown in this paper, the scenarios, categories, and statements enable research into how humans provide emotional support, and the development and evaluation of emotional support algorithms. We expect that the validated statements in particular will be useful for others' research, as will the method used here to obtain them.

Next, in Phase 2 of the research, three studies inspired the development of and evaluated the effectiveness of an algorithm that generates emotional support tailored to a specific stressor. Study 4 showed that people use differing categories of emotional support when supporting others, depending on the stressor which that person is experiencing. Interestingly, we also found that despite all emotional support statements having been reliably classified into categories some statements within each category were more favourably used. From the results, three emotional support algorithms were developed. These algorithms were then refined by Study 5 resulting in the development of one 'optimized' algorithm. This process revealed four unique emotional support strategies for providing support for specific stressors. Each identified emotional support category was used by at least one of these strategies. Study 6 evaluated our algorithm from the point of view of the receiver of the support. The results suggested that the algorithm performed very well for Isolation, Interruption and Physical Demand, as the tailored support generated for those stressors was rated as most effective for those stressors. The messages designed for Emotional Demand, Mental Demand and Frustration all

Table 14
Summary of studies.

Study	Participants	Goal	Results
1	30	Validation of scenarios that exhibit specific stressors	7 validated scenarios depicting Frustration, Isolation, Interruption, Temporal Demand, Emotional Demand, Mental Demand, Physical Demand (Table 2)
2	20	Generation of emotional support statements	85 unique support statements (Table 5 shows 53 of these)
3	40	Validation of emotional support statements' categorization	5 emotional support categories: Direct Action, Emotional Advice, Emotional Reflection, Praise and Reassurance (Fig. 4 shows definitions) 53 statements validated as belonging to these categories (Table 5). 20 statements – 4 best per category – used in next study
4	100	Generation of initial emotional support algorithms	People tailor use of support categories to stressors Data showing what categories (Fig. 7), combinations (Table 6), and ordering are used per stressor Data showing frequency of usage of individual support messages from the categories per stressor (Table 8) 3 tailoring algorithms and Dice scores showing how well they fit the data (see Table 7)
5	100	Evaluation and refinement of algorithms	Data showing perceived goodness of the 3 algorithms' messages for the stressor they were generated for (Table 9) Algorithm 2 performed best, and Algorithm 1 worst All algorithms performed poorly on Temporal Demand Algorithms 2 and 3 performed quite well on other stressors One optimized algorithm (see Table 10)
6	100	Evaluation of final algorithm	Data showing perceived goodness of all messages generated by the final algorithm for all stressors and of all support strategies used (Fig. 10 and Table 13) The effectiveness of support messages and strategies depends on the stressor (Fig. 10, Tables 12 and 13) The algorithm performed well on all stressors except Temporal Demand, with messages generated for the stressor being more effective than or statistically equally effective as other messages (Fig. 10 and Table 12) The message generated for Physical Demand (ER+PR) performed well for all stressors (Fig. 10), and was better for Temporal Demand than the algorithm's message PR+RE and ER+PR performed equally well on any stressor and better than PR+DA, indicating that RE and ER are perceived as equally effective and as more effective than DA ER+EA performed reasonably well on all stressors PR+DA was relatively ineffective except for Mental Demand, one of the two stressors for which it was designed (the other being Temporal Demand), indicating that people may only want to receive DA when the stress is down to their abilities

performed adequately, with no other messages being rated significantly better for those stressors. Only the support designed for Temporal Demand performed badly, with messages designed for other stressors being rated significantly higher for this scenario. This was not unexpected as the initial ratings of the messages designed for Temporal Demand during the development of our algorithm received a lower rating of effectiveness than the messages for other stressors. The results also showed that certain emotional support strategies were better received than others with the emotional support strategy consisting of 'Emotional Reflection' and 'Praise' performing well for all stressors, while the strategy consisting of 'Praise' and 'Directed Action' performed poorly for all scenarios except for that describing mental demand.

In Phase 2, we learned that emotional support needs to be tailored to the stressor experienced, and that crowd-sourcing can be used to (1) inspire emotional support algorithms, (2) evaluate the perceived effectiveness of emotional support algorithms. We learned about the relative effectiveness of different strategies for different stressors. We also learned that even when using emotional support categories, one still needs to be careful with the design of individual messages within a category. We expect that the method we followed here can be of use to other researchers, and that the resulting algorithm (with an adaptation for Temporal Demand where it did not perform well) will be implemented in virtual agents and further investigated. Our recommendations to practitioners can be seen in Table 15.

5.2. Directions for future work

We validated scenarios which each described one stressor. Research is needed to design scenarios that describe combinations of stressors, as multiple stressors can be present in a situation simultaneously (as also indicated in discussions we have had with CFRs). Such scenarios would enable an investigation into whether the emotional support that people provide or wish to receive varies depending on combinations of stressors, and the development of a more sophisticated emotional support algorithm.

As discussed above, there is a possible limitation due to the use of only one scenario to depict a stressor, and it would be good to develop additional scenarios, to be able to investigate whether the wordings of the scenarios have impacted the strategies used.

We validated a corpus of categorised emotional support statements.This corpus could be further populated by future research allowing for a greater depth and richness of statements from which the algorithm can generate emotional support. This may also provide the opportunity to apply techniques such as Natural Language Generation allowing the algorithm to 'go beyond' the restrictions of concatenating statements together to generate support.

The algorithm produced adapts to stressors, but not yet to user characteristics. We want to investigate how user characteristics such as personality, gender, age, and cultural background may impact what support people would provide, would like to receive, and how they would react to support provided by a virtual agent. Tailoring support to these characteristics may lead to a more effective algorithm. For example, when we investigated the use of individual support categories (see Section 4.1.2), we found that *praise* was not the most appropriate strategy for any of the scenarios. However, praise is related to Esteem support (see Section 3.4), and it can be hypothesized that esteem support is more needed for people who lack self-esteem. Within the domain of e-learning, there has been some work showing that emotional support needs adapting to learner personality (Dennis et al., 2015).

We have not yet implemented the algorithm into a virtual agent. Before the studies presented here, we had already produced

Table 15
Recommendations to practitioners.

Stressor	Categories	Message
Stressor unknown	ER + PR	I know this is hard but you are doing a great job
Temporal Demand	ER + PR	I know this is hard but you are doing a great job
Physical Demand	ER + PR	I know this is hard but you are doing a great job
Mental Demand	PR + DA	You are doing a great job but pay attention to detail
Emotional Demand	ER + EA	I know this is hard but stay positive
Interruption	ER + PR	I understand that this is frustrating however good job at keeping your cool
Frustration	ER + EA	I understand that this is frustrating but be glad that you can help
Isolation	PR + RE	You are doing a great job, it is going to be fine

a virtual agent that attempted to alleviate user stress (see Kindness et al., 2013) and measured its impact on stress through sensors and self-reporting questionnaires. While performing that work, we realized that we needed to more thoroughly understand how best to provide support which formed the rationale for this research. The next step would be to implement our algorithm into this virtual agent.

Our studies did not evaluate how effective the emotional support was considered to be by participants when they were experiencing actual stress. In future research we could seek to investigate how effective the support was considered to be in a stressful situation, using a methodology similar to the one we used in Kindness et al. (2013) or immersing participants in a controlled stressful scenario in a virtual reality setting. To achieve this, task based scenarios would need to be developed that could be undertaken within a controlled environment, which would induce a specific stressor. Additionally, a longitudinal study could be carried out with the agent providing emotional support over a longer period.

The relationship between the virtual agent and the user may impact the effectiveness of the emotional support. In our previous research we portrayed the virtual agent as having a relationship that is *one-up* to the participant, i.e. the agent was portrayed as an expert. Alternatively, the relationship could be portrayed as *one-across*, i.e. the agent being portrayed as a peer. The effect of this on emotional support needs to be investigated.

The agent's appearance is also likely to impact on the effectiveness of its emotional support. The development of intelligent agents that interact with humans in a human like manner, with dynamic and rich interactions such as appearance and movement has been the focus of innumerable studies (e.g. Cassell et al., 2000; Nass et al., 2000; Badler et al., 2002; Poggi et al., 2005). Research in to how virtual agent can effectively portray emotions (cf. Bartneck et al., 2004; Koda and Ishida, 2006) may also reinforce the emotional support which a virtual agent is trying to convey. As well as appearance, there are several experimental approaches and aspects of interaction surrounding embodied agents that could be explored (cf. Dehn and Van Mulken, 2000).

People may be reluctant to believe that they can receive emotional support from a computer (as raised in our discussions with CFRs), and an investigation is needed into how best to provide support to somebody while they are in a stressful situation (e.g., audibly or visually), and how to best incorporate this in systems used.

Finally, the virtual agent needs to be able to determine the affective state of the user and the stressor experienced. Substantial advances have been made within the affective computing research field allowing automatic detection of a person's emotional states, for example based on posture (Kleinsmith and Bianchi-Berthouze, 2007; De Silva and Bianchi-Berthouze, 2004) or facial and vocal expressions (Suzuki and Naitoh, 2003; Meng and Bianchi-Berthouze, 2014). However, being able to determine these within a real-world environment and determining the stressor experienced remains a challenge.

Acknowledgements

This work was funded by the RCUK Digital Economy award to the dot.rural Digital Economy Hub, University of Aberdeen; award reference: EP/G066051/1. The dataset used by this paper can be acquired by emailing the first author. We thank Matt Dennis, Kirsten A. Smith and Michael Gibson for their contributions to the research.

References

APA, 2013. The American Psychological Association – Types of stress, http://www.apa.org/helpcenter/stress-kinds.aspx.
Badler, N., Allbeck, J., Zhao, L., Byun, M., 2002. Representing and parameterizing agent behaviors. In: Proceedings of Computer Animation, 2002, pp. 133–143.
Barbee, A.P., Cunningham, M.R., Winstead, B.A., Derlega, V.J., Gulley, M.R., Yankeelov, P.A., Druen, P.B., 1993. Effects of gender role expectations on the social support process. J. Soc. Issues 49 (3), 175–190.
Bartneck, C., Takahashi, T., Katagiri, Y., Dolech, D., 2004. Cross-cultural study of expressive avatars. In: Proceedings of the Social Intelligence Design, pp. 21–27.
Baylor, A.L., Warren, D., Park, S., Shen, E., Perez, R., 2005. The impact of frustration-mitigating messages delivered by an interface agent. In: Artificial Intelligence in Education. IOS Press, Amsterdam, pp. 73–79.
Billings, A.G., Moos, R.H., 1981. The role of coping responses and social resources in attenuating the stress of life events. J. Behav. Med. 4 (2), 139–157.
Brave, S., Nass, C., Hutchinson, K., 2005. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. Int. J. Human–Comput. Stud. 62, 161–178.
Burleson, B.R., Kunkel, A.W., 1996. Handbook of social support and the family. In: The Socialization of Emotional Support Skills in Childhood. Springer, New York, pp. 105–140.
Burleson, W., Picard, R.W., 2007. Gender-specific approaches to developing emotionally intelligent learning companions. IEEE Intell. Syst. 22 (4), 62–69.
Burleson, B.R., 1985. The production of comforting messages: social-cognitive foundations. J. Lang. Soc. Psychol. 4, 253–273.
Carver, C.S., Scheier, M.F., Weintraub, J.K., 1989. Assessing coping strategies: a theoretically based approach. J. Personal. Soc. Psychol. 56 (2), 267.
Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsson, H., Yan, H., 2000. Designing embodied conversational agents. Embodied Conversat. Agents, 29.
Cutrona, C.E., Russell, D.W., 1990. Type of social support and specific stress: toward a theory of optimal matching. In: Social Support: An Interactional View. John Wiley and Sons, Oxford, pp. 319–366.
Cutrona, C.E., Suhr, J.A., 1992. Controllability of stressful events and satisfaction with spouse support behaviors. Commun. Res. 19 (2), 154–174.
De Silva, P.R., Bianchi-Berthouze, N., 2004. Modeling human affective postures: an information theoretic characterization of posture features. Comput. Anim. Virtual Worlds 15 (3–4), 269–276.
Deemter, K.v., Gatt, A., Sluis, I.v.d., Power, R., 2012. Generation of referring expressions: assessing the incremental algorithm. Cogn. Sci., 799–836.
Dehn, D.M., Van Mulken, S., 2000. The impact of animated interface agents: a review of empirical research. Int. J. Human–Computer Stud. 52, 1–22.
Dennis, M., Kindness, P., Masthoff, J., Mellish, C., Smith, K., 2013. Towards effective emotional support for community first responders experiencing stress. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII). IEEE, Geneva, pp. 763–768.
Dennis, M., Masthoff, J., Mellish, C., 2013. Does Learner Conscientiousness Matter When Generating Emotional Support in Feedback? In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII). IEEE, Geneva, pp. 209–214.
Dennis, M., Masthoff, J., Mellish, C., 2015. Adapting progress feedback and emotional support to learner personality. Int. J. Artif. Intell. Educ., 1560–4306, 1–55 ISSN.
Fleiss, J.L., 1981. Statistical Methods for Rates and Proportions, second ed. John Wiley & Sons, New York.
Fogg, B.J., Nass, C., 1997. Silicon sycophants: the effects of computers that flatter. Int. J. Hum–Comput. Stud. 46, 551–561.
Folkman, S., Lazarus, R.S., 1980. An analysis of coping in a middle-aged community sample. J. Health Soc. Behav., 219–239.
Gilliland, A.L., 2011. After praise and encouragement: emotional support strategies used by birth doulas in the USA and Canada. Midwifery 27 (4), 525–531.
Goldsmith, D.J., 2004. Communicating Social Support. Cambridge University Press, New York.
Gratch, J., Marsella, S., 2001. Tears and fears: modeling emotions and emotional behaviors in synthetic agents. In: Proceedings of the Fifth International

Conference on Autonomous Agents, AGENTS '01. ACM, New York, NY, USA, pp. 278–285.
Gross, J.J., 1998. The emerging field of emotion regulation: an integrative review. Rev. Gen. Psychol. 2 (3), 271.
Gross, J.J., 2002. Emotion regulation: affective, cognitive, and social consequences. Psychophysiology 39 (3), 281–291.
Hart, S.G., 2006. NASA-task load index (NASA-TLX); 20 years later. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting 50, pp. 904–908.
Hone, K., 2006. Empathic agents to reduce user frustration: the effects of varying agent characteristics. Interact. Comput. 18, 227–245.
Johnson, D., Gardner, J., Wiles, J., 2004. Experience as a moderator of the media equation: the impact of flattery and praise. Int. J. Hum.–Comput. Stud. 61 (3), 237–258.
Jones, S.M., Burleson, B.R., 1997. The impact of situational variables on Helpers' perceptions of comforting messages an attributional analysis. Commun. Res. 24, 530–555.
Kindness, P., Mellish, C., Masthoff, J., 2013. How virtual teammate support types affect stress. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII). IEEE, Geneva, pp. 300–305.
Kindness, P., Fitzpatrick, D., Mellish, C., Masthoff, J., O'Meara, P., McEwan, M., 2014. An Insight into the demands and stressors experienced by community first responders. J. Paramed. Pract. 6, 362–369.
Klein, J., Moon, Y., Picard, R.W., 2002. This computer responds to user frustration: theory, design, and results. Interact. Comput. 14 (2), 119–140.
Kleinsmith, A., Bianchi-Berthouze, N., 2007. Recognizing affective dimensions from body posture. Affect. Comput. Intell. Interact., 48–58.
Koda, T., Ishida, T., 2006. Cross-cultural study of avatar expression interpretations. In: Proceedings of the International Symposium on Applications on Internet. IEEE Computer Society, Washington, pp. 130–136.
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174.
Lazarus, R.S., Folkman, S., 1984. Stress, Appraisal, and Coping. Springer Publishing Company, New York.
Lee, E.-J., 2008. Flattery may get computers somewhere, sometimes: the moderating role of output modality, computer gender, and user gender. Int. J. Hum.–Comput. Stud. 66 (11), 789–800. http://dx.doi.org/10.1016/j.ijhcs.2008.07.009, ISSN 1071-5819 ⟨http://www.sciencedirect.com/science/article/pii/S107158190800092X⟩.
Lehman, D.R., Hemphill, K.J., 1990. Recipients' perceptions of support attempts and attributions for support attempts that fail. J. Soc. Person. Relationsh. 7, 563–574.
Lehman, D.R., Ellard, J.H., Wortman, C.B., 1986. Social support for the bereaved: recipients and providers perspectives on what is helpful. J. Consult. Clin. Psychol. 54, 438.
Masthoff, J., 1997. An Agent-based Interactive Instruction System (Ph.D. thesis). Technical University, Eindhoven, The Netherlands.
Masthoff, J., 2006. The user as wizard: a method for early involvement in the design and evaluation of adaptive systems. In: Workshop on User-Centred Design and Evaluation of Adaptive Systems, pp. 460–469.
McCrae, R.R., Costa, P.T., 1986. Personality, coping, and coping effectiveness in an adult sample. J. Personal. 54 (2), 385–404.
Meng, H., Bianchi-Berthouze, N., 2014. Affective state level recognition in naturalistic facial and vocal expressions. IEEE Trans. Cybern. 44 (3), 315–328.
Meyer, D.K., Turner, J.C., 2002. Discovering emotion in classroom motivation research. Educ. Psychol. 37 (2), 107–114.
MIME, Managing Information in Medical Emergencies project, ⟨http://www.dotrural.ac.uk/mime⟩, 2013.
MT, 2013. Amazon Mechanical Turk, ⟨http://www.mturk.com⟩.
Nash, C., Steurer, J., Tauber, E.R., 1994. Computers are social actors. In: Proceedings of SIGCHI, pp. 74–78.
Nass, C., Isbister, K., Lee, E.-j., 2000. Truth is beauty: researching embodied conversational agents. Embodied Conversat. Agents, 374–402.
Nguyen, H., Masthoff, J., 2009. Designing empathic computers: the effect of multimodal empathic feedback using animated agent. In: Proceedings of the 4th International Conference on Persuasive Technology. ACM, New York, p. 7.
Paiva, A., Dias, J., Sobral, D., Aylett, R., Sobreperez, P., Woods, S., Zoll, C., Hall, L., 2004. Caring for agents and agents that care: building empathic relations with synthetic agents. In: AAMAS, pp. 194–201.
Paraboni, I., Van Deemter, K., Masthoff, J., 2007. Generating referring expressions: making referents easy to identify. Comput. Linguist. 33 (2), 229–254.
Paramythis, A., Weibelzahl, S., Masthoff, J., 2010. Layered evaluation of interactive adaptive systems: framework and formative methods. User Model. User-Adapt. Interact. 20 (5), 383–453.
Picard, R.W., Klein, J., 2002. Computers that recognise and respond to user emotion: theoretical and practical implications. Interact. Comput. 14, 141–169.
Pierce, G.R., Sarason, I.G., Sarason, B.R., 1991. General and relationship-based perceptions of social support: are two constructs better than one? J. Personal. Soc. Psychol. 61, 1028.
Poggi, I., Pelachaud, C., de Rosis, F., Carofiglio, V., De Carolis, B., 2005. Greta: a believable embodied conversational agent. In: Multimodal Intelligent Information Presentation. Springer, Dordrecht, pp. 3–25.
Prendinger, H., Ishizuka, M., 2004. Life-like Characters: Tools, Affective Functions And Applications. Springer, Berlin.
Prendinger, H., Ishizuka, M., 2005. The empathic companion: a character-based interface that addresses users' affective states. Appl. Artif. Intell. 19, 267–285.
Prendinger, H., Mayer, S., Mori, J., Ishizuka, M., 2003. Persona effect revisited. Using Bio-signals to measure and reflect the impact of character-based interfaces. In: Intelligent Virtual Agents. Springer, Berlin, Heidelberg, pp. 283–291.
Randolph, J.J., 2005. Free-marginal multirater kappa: an alternative to Fleiss' fixed-marginal multirater kappa. In: Joensuu University Learning and Instruction Symposium, 2005.
Robison, J., McQuiggan, S., Lester, J., 2010. Developing empirically based student personality profiles for affective feedback models. In: ITS. Springer, Berlin, Heidelberg, ISBN 978-3-642-13387-9, pp. 285–295.
Rook, K., Underwood, L., 2000. Social Support measurement and intervention: comments and future directions. In: Social Support Measurement and Interventions: A Guide for Health and Social Scientists. Oxford University Press, New York, pp. 311–334.
Roth, S., Cohen, L.J., 1986. Approach, avoidance, and coping with stress. Am. Psychol. 41 (7), 813.
Scheffe, H., 1953. A method for judging all contrasts in the analysis of variance. Biometrika 40 (1–2), 87–110.
Selye, H., 1956. The Stress of Life. McGraw-Hill, New York.
Smith, K.A., Masthoff, J., Tintarev, N., Moncur, W., 2014. The development and evaluation of an emotional support algorithm for carers. Intell. Artif. 8 (2), 181–196.
Suzuki, K., Naitoh, K., 2003. Brief report: useful information for face perception is described with FACS. J. Nonverbal Behav. 27, 43–55.
Taylor, W.L., 1953. Cloze procedure: a new tool for measuring readability. Journal. Q. 30, 415–433.
van Deemter, K., 2016. Computational Models of Referring: A Study in Cognitive Science. MIT Press, Cambridge, MA.
van der Zwaan, J.M., Dignum, V., Jonker, C.M., 2012. A Conversation Model Enabling Intelligent Agents to Give Emotional Support. Springer, Berlin, Heidelberg.
Xiao, Y., Wang, Z., Wang, M., Lan, Y., 2005. The appraisal of reliability and validity of subjective workload assessment technique and NASA-task load index. Zhonghua lao dong wei sheng zhi ye bing za zhi (Chinese journal of industrial hygiene and occupational diseases) 23, 178–181.
Young, P.M., Gibson, A.S.C., Partington, E., Partington, S., Wetherell, M., 2013. Anxiety, stress, and perceived workload during the command and control of computer-simulated fire service training environments. Int. J. Emerg. Serv. 2, 119–130.
Zeidner, M., Endler, N.S., 1996. Handbook of Coping: Theory, Research, Applications, vol. 195. John Wiley & Sons, New York.