Undefined 1 (2014) 1–5
IOS Press

The Development and Evaluation of an Emotional Support Algorithm for Carers

Kirsten A. Smith [a,*], Judith Masthoff [a] Nava Tintarev [a] and Wendy Moncur [b]
[a] *Computing Science, University of Aberdeen, Aberdeen, United Kingdom*
E-mail: {r01kas12, j.masthoff, n.tintarev}@abdn.ac.uk
[b] *University of Dundee, Dundee, United Kingdom*
E-mail: w.moncur@dundee.ac.uk

Abstract. Carers - people who provide regular support for a friend or relative who could not manage without them - frequently report high levels of stress. Good emotional support could help relieve this stress. This study uses seven scenarios that depict different types of stress and acquires emotional support messages for them. We then categorize and evaluate the emotional support for different types of stress. We found that telling the carer they are appreciated and offering support are the best types of emotional support. Additionally, we found that how well a supporter sympathises with a situation affects the type of support they consider suitable. We describe and evaluate an algorithm that selects different categories of support to be used by an intelligent virtual agent to provide emotional support to carers experiencing different types of stress.

Keywords: agents, e-health, emotional support, stress, carers

1. Introduction

Carers are people who provide regular support for a friend or relative who could not manage without them, without formal payment. They save the UK economy £119 billion per year [7], but frequently report high levels of stress [1,25]. Good emotional support has been found to reduce negative affect [6,19] and could help relieve this stress. However, carers have less time to maintain social relationships due to their caring commitments and thus are less able to obtain emotional support from their personal social network (e.g. friends and family).

One solution is to create an intelligent virtual agent (IVA) that can itself offer sensitive, suitable emotional support at times of stress. Emotional support agents that react to affect (e.g. [23]) have been used to improve learning outcomes [26], increase interaction time with a system [15], decrease stress levels [21] and reduce negative affect [19]. Dennis et al. [9] created a corpus of empathetic support statements for an agent to use to support community first responders experiencing different kinds of stress. However, there has been no investigation into developing an intelligent agent to provide emotional support to carers who have different needs to community first responders. Carers experience longer periods of lower-level stress which they cannot escape from, in contrast to community first responders who experience more short-term stress.

The aim of emotional support can be seen as positively modifying the emotional response to a situation. Gross[12] describes a model by which the emotional response to an event is modulated by the individual's Situation Selection, Modification, Attention, Cognitive Change and Modulation. Emotional support provided after the event can encourage cognitive change e.g. reappraising the stressful situation in a more positive light. It could also offer advice on how to positively modify future situations and discourage maladaptive coping strategies.

*Corresponding author. E-mail: r01kas12@abdn.ac.uk

Burleson [8] found that good emotional support messages are person-centred: they acknowledge and elaborate on another's feelings (e.g. 'I understand you're frustrated, it must be really hard!'). Low person-centered messages criticise feelings and are directive (e.g. 'There's no point worrying, just get it done'). Similarly, Barbee et al [4] describe a model (see Figure 1) where support is either emotion- or problem- focused and either approach- or avoid- based. Approach-emotion 'Solace' strategies elicit positive emotion and express closeness; approach-problem 'Solve' strategies attempt to help solve the problem; avoid-emotion 'Escape' strategies discourage negative emotion and distract the person and avoid-problem 'Dismiss' strategies downplay the significance of the problem. Approach-based support was found to be the most effective, especially 'Solace' [5] (which is high person-centred). In this paper we break this down further and identify categories of high person-centred emotional support that are effective when given to carers experiencing different types of stress.

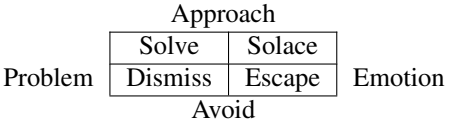

Fig. 1. Interactive Coping Behaviours [4]

To obtain a corpus of emotional support messages, we created scenarios depicting stressors that carers might experience and asked participants to provide emotional support messages for them. Subsequently, we categorised the messages into support types using a card sort task and then evaluated how suitable participants thought the categories were at providing emotional support for carers in our care scenarios. This lead us to develop an algorithm for emotional support, which we then refined and evaluated.

Our study was thus broken down into 5 experiments: Scenario Validation, Emotional Support Acquisition, Emotional Support Categorization, Emotional Support Algorithm Creation and Algorithm Evaluation.

2. Methodology

Similarly to [9], we used the User as Wizard method [18] to obtain emotional support messages from participants. This method places the participants in the role of the virtual agent, providing support for the carer in the stressful situation. Following [10]'s methodology, we decided to develop scenarios that would depict the carer in the stressful situation, rather than simply telling the user the stressor e.g. 'Bob is feeling frustrated'. We believe that developing believable, empathetic scenarios both makes the users more likely to empathise and give a richer, more person-centred response and makes our final results more generalisable to real-world care scenarios.

For all experiments described in this paper, participants were recruited from Amazon's Mechanical Turk (a crowd-sourcing tool) [20]. This allowed us (*the requester*) to create short tasks (*HITs*) which participants (*workers*) were paid $0.50 to complete. Participants were recruited from the US only and were required to have a Mechanical Turk acceptance rate of greater than 90% (at least 90% of their HITs are considered of good quality by other requesters). They were also required to correctly complete a Cloze Test [24] for English fluency so that we could ensure the participants were able to read and understand the task. Participants were then presented with the task (for example, see Figure 2).

3. Experiment One: Scenario Validation

This experiment validated scenarios to ensure that they depicted a particular type of stressful situation a carer may experience. These scenarios will be used in future experiments to create a context for support messages to carers. This will enable us to offer carers support that was suitable for the situation they had experienced.

Table 1
Stressors adapted from the NASA-TLX [13] by [9]

Code	Name	Description- The scenario portrays...
ED	Emotional Demand	emotional demand, such as feelings of empathy with another.
FR	Frustration	a feeling of frustration or annoyance with the activity.
IN	Interruption	the stress of interruptions during an activity.
IS	Isolation	loneliness and isolation.
MD	Mental Demand	a mentally taxing activity in which the carer needs to think.
PD	Physical Demand	stress caused by physical demands on the carer.
TD	Temporal Demand	a sense of time pressure.

Table 2
Scenarios validated for each stressor. κ=Free-marginal kappa 1.0=excellent agreement, 0.7=good agreement, 0.4=moderate agreement

Stressor*	Scenario	κ
IN	Susan is John's carer. Today Susan needed to get John ready for bed, but people kept phoning her.	0.71
IS	Fiona is Fred's carer. Fred spends most of the day asleep. Today Fiona was alone all day and no home carers were scheduled to visit.	1.00
MD	Martin is Julia's carer. Today Martin had to carry out minor medical tests. The tests are not dangerous if he does them wrong but the procedure is complex and requires concentration.	0.85
PD	Carol is Max's carer. Today Carol moved heavy furniture and boxes from Max's upstairs bedroom to his new bedroom downstairs.	0.92
TD	Ben is Samantha's carer. Today Ben had to drop Samantha off at the doctors at 4.30pm, collect her prescription from the pharmacy at the other side of town before it closed and collect some groceries before collecting her at 5pm.	0.92
ED	Andrea is Gary's carer. Today Gary was confused and very upset and Andrea comforted him.	0.65
FR	Harry is Diane's carer. Today Harry wanted to drop Diane off at the day care center so he could have some free time, but the center was closed.	0.39

*See Table 3 for all abbreviations

Table 3
Abbreviations for Scenarios & Categories

Support Categories		Scenarios	
APP	Appreciated	MD	Mental Demand
SUP	Supported	TD	Temporal Demand
EMP	Empathy	PD	Physical Demand
CON	Consolation	FR	Frustration
PRA	Practical Advice	IN	Interruption
EMO	Emotional Advice	IS	Isolation
ENC	Encouragement	ED	Emotional Demand
DES	Deserving		
BLA	Blameless		
PRS	Praise		

3.1. Design

A within-subject design was used: each participant considered seven scenarios, each created to depict a stressor (see Table 1). Four rounds of testing were implemented so that scenarios that were not well classified could be replaced or adjusted.

3.2. Participants

Each round had 30 participants (no participant took part in more than 1 round). There were 120 participants in total; 53% of participants were male, 47% female; 20% were aged 16-25, 50% were 26-40, 28% were 41-65, 1% were over 65 and 1% did not disclose their age.

3.3. Materials

Sixteen scenarios were tested in total, see Table 2 for the final validated set. Scenarios describe a carer and their caree (the person they care for. Each scenario was intended to reflect one of 7 key stressors (see Table 1), adapted from the NASA-Task Load Index [13] by [9].

3.4. Procedure

Participants were presented with seven scenarios in random order and were asked which stressor they thought that each depicted (from the set of 7 stressors; see Table 1).

3.5. Results

A free-marginal kappa (κ) [22] was used to assess agreement between participants; 1 indicates unanimous agreement, 0.7 excellent agreement and 0.4 moderate agreement. The validated scenario set is shown in Table 2.

Frustration scenarios were not well classified - we tested 7 different frustration scenarios and selected the best one with a kappa of 0.39. It might be that frustration scenarios commonly involve multiple types of stressor; it is included in our analysis as a baseline non-specific stressful situation.

4. Experiment Two: Emotional Support Acquisition

Once we had scenarios that depicted stressors that carers might face, we ran an experiment to gather a corpus of emotional support statements that people might provide to carers in different situations.

4.1. Design

A within-subject design was used: each participant considered seven scenarios.

4.2. Participants

There were 31 participants. 48% were male, 52% female; 26% were aged 16-25, 29% were 26-40, 45% were 41-65.

4.3. Materials

The seven validated scenarios from Experiment One were used, see Table 2.

4.4. Procedure

The experiment began by explaining the concept of a carer to participants (i.e. explaining that we meant informal carers as opposed to nurses or other professionals). Participants were then given examples of emotional support messages (obtained by [11] previously for learners) to illustrate what we meant by emotional support.

Participants were presented with each of the scenarios in turn and asked to provide 3 short messages of emotional support to the carer. Afterwards, participants were given the opportunity to provide comments.

4.5. Results

A corpus of 651 support messages was produced, such as "I am here for you" and "You are a good person"(see Table 7 for more examples).

5. Experiment Three: Emotional Support Categorization

The aim of this experiment was to investigate which emotional support messages out of a set of 114 (see examples in Table 7) could be reliably identified as belonging to particular emotional support categories (see Table 5).

Table 4
Initial categories arising from the card sort task and their final categories. Italics indicate subcategories.

Code	Initial Categories	Final Categories
0	Rubbish	*Not included*
1	Caree benefits	Appreciation
2	Offers of help	Supported
2.1	*listening*	Supported
3	Empathy	Empathy
4	Consolation	Consolation
5	Practical general advice	Practical advice
5.1	*Practical situational advice*	*Not included*
6	Emotional advice	Emotional advice
7	Encouragement	Encouragement
8	Future Success Assertions	Encouragement
9	Carer deserves reward	Deserving
10	Carer not to blame	Blameless
11	General support	*Not included*
12	Praise	Praise
12.1	*Good carer*	Appreciation
12.2	*Efficient/cope well*	Praise
12.3	*Good effort*	Praise
12.4	*hard work*	Praise
12.5	*patient*	Praise
12.6	*strong*	Praise
12.7	*morally right*	Praise
12.8	*Super*	Praise
12.9	*Good person*	Praise
12.11	*Capable*	Praise
12.12	*Thanks*	Praise
12.13	*Good intentions*	Praise
12.14	*Caring*	Praise
12.15	*skilled*	Praise
12.16	*Good job*	Praise
12.17	*misc praise*	Praise
12.18	*dedicated*	Praise
13	other	*Not included*

5.1. Design

The experiment was run in two rounds: to reduce workload on participants, the 114 messages were split into two sets of 57, with participants only considering one of the sets. A within-subject design was used: each participant considered 57 emotional support messages.

5.2. Participants

There were 55 participants in total. 51% were male, 49% female; 20% were aged 16-25, 45% were 26-40, 31% were 41-65 and 4% were over 65.

5.3. Materials

We used ten categories of emotional support (see Table 5) and 144 unique emotional support messages (see Table 7 for examples).

Emotional support categories were derived from an open card sort task on the corpus of emotional support messages from Experiment Two. Each message was written on a separate piece of paper and laid out on a table. Messages were then sorted into groups based on similarity and labels assigned to these groups (see Table 4). On discussion, some categories were merged with other categories and the subgroups of 'Praise' were not used (there were too few of each subgroup to be meaningful in analysis; however these should be explored in future work). This resulted in the final set of ten categories.

Emotional support messages were derived from the corpus of emotional support messages from Experiment Two. We removed duplicate or semantically similar messages from the set (e.g. 'Breathe' and 'Take a deep breath') and messages including scenario-specific information (e.g. 'Just focus on Zack, and ignore the phone') so that the message set is generalisable. Finally, we removed any names or genders in the messages and replaced them with a marker so that the correct names could be inserted for later use. This left us with the 114 unique support messages.

5.4. Procedure

We presented participants with each message from the set, one by one in a random order, and asked them to select the category from Table 5 that they thought the message belonged to. If they felt the message belonged to none of the categories, they could select "other". Participants could provide free-text comments after the experiment had finished.

5.5. Results

A free-marginal kappa [22] was used to assess agreement between participants. 63 statements had a κ>0.4 (see Table 5 for breakdown of categories and Table 7 for the messages).

6. Experiment Four: Emotional Support Algorithm Creation

The next step was to evaluate how suitable different categories of emotional support are in the 7 stressful scenarios in order to produce an algorithm for the intelligent virtual agent to use.

6.1. Design

We used a between-subject design: each participant considered one randomly assigned scenario. To reduce the workload on the participant, each participant was only asked to rate a set of 20-30 messages for this scenario. They could then choose to repeat this up to four times with different message sets.

The independent variables were:

- Message (63 levels): The emotional support message under consideration.
- Original Scenario (7): The scenario the message was generated for in Experiment Two.
- Presented Scenario (7): The scenario presented to the participant in this experiment (which may be a different scenario than the one the message was generated for).
- Message Category (10): The emotional support category the message was reliably classified as in Experiment Three.

The dependent variable was Suitability, expressed by 4 measures: appropriateness, effectiveness, helpfulness and sensitivity (see Figure 2), all measured on a Likert scale from 1 (worst) to 9 (best). These scales were used by [14] and were found to be internally consistent, measuring the single factor 'Suitability'.

6.2. Hypotheses

- **H1**: Different Message Categories will be rated higher in different Presented Scenarios
- **H2**: Empathetic, person-centered support Categories will be rated highest, in line with [8]. Message categories Appreciated, Praise, Supported, Empathy, Encouragement and Deserving are broadly equivalent to high empathetic person-centred support and should thus be judged as higher quality emotional support.
- **H3**: Messages will be rated as most suitable when presented with the Original Scenario.

6.3. Participants

There were 116 participants. 59% of participants were male, 41% female; 24% were aged 16-25, 50% were 26-40, 25% were 41-65 and 1% did not disclose their age.

Table 5
Support Message Categories with Number of Categorised Messages

Code	Category	Description	No κ>0.4
APP*	Appreciated	Reminds the carer that what they are doing is beneficial to someone else.	8
PRS*	Praise	Praises the carer, making them feel good about themselves.	20
SUP*	Supported	Offers to do something to help the carer.	6
EMP*	Empathy	Acknowledges how the carer is feeling.	6
DES*	Deserving	Suggests that the carer should be rewarded.	1
ENC*	Encouragement	Asserts that the carer is capable, encouraging them to do or continue something.	4
BLA	Blameless	Reassures the carer that the situation is not their fault.	2
CON	Consolation	Suggests a positive interpretation of the situation.	1
PRA	Practical Advice	Suggests what to do or the manner in which to do it.	12
EMO	Emotional Advice	Suggests how the carer should feel.	3

*Person-centred support categories [8]

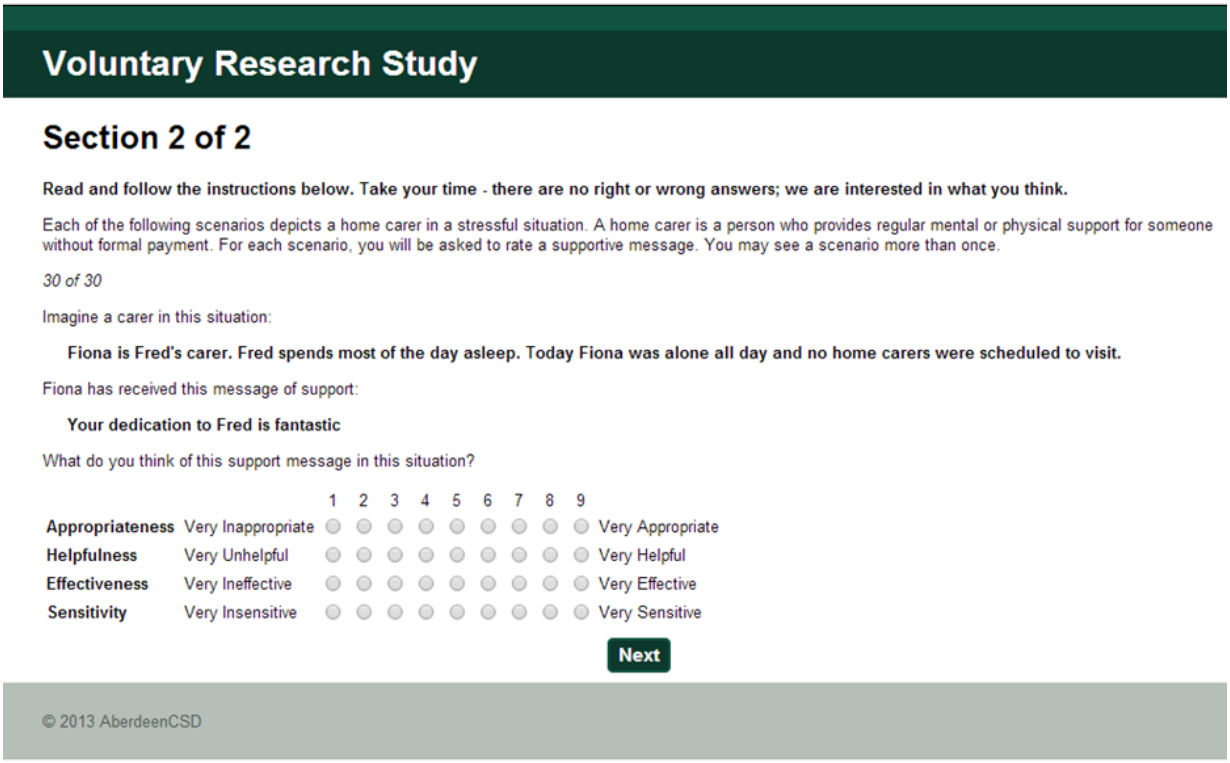

Fig. 2. Screenshot of Experiment 4

6.4. Materials

We used the seven validated scenarios from Experiment One (see Table 2) and the 63 reliably categorized support messages from Experiment Three (see Table 7).

6.5. Procedure

Participants were presented with a scenario (Presented Scenario) and a randomized emotional support message (Message). They were asked to rate the message on the four Suitability sub-scales.

6.6. Results

Each participant rated between 14 and 111 scenario/message pairs (median 21)[1]. Each scenario/message pair was rated at least five times. No single participant rated all scenario/message pairs or rated a scenario/message pair more than once. We thus chose to analyse this as a between-subjects design.

6.6.1. Manipulation Check.

To ensure that the average of the 4 rating types was appropriate for statistical analysis, we conducted a manipulation check. We found that the 4-item measures (appropriateness, helpfulness, effectiveness and sensitivity) were internally consistent (Cronbach's alpha=0.84). A Principal Component Analysis confirmed that the 4 items measured a single factor 'Suitability' for all of the categories (eigenvalues ranged from 3.23-3.71) and scenarios (eigenvalues 3.45-3.59).

6.6.2. Effects of Message Category × Presented Scenario.

A 7×10 2-way ANOVA was performed on presented scenario and support category on suitability. There were significant effects for message category ($F(9,11851)=205.88$, $p<0.001$), presented scenario ($F(6,11851)=17.42$, $p<0.001$) and for the interaction ($F(54, 11851)=10.51$, $p<0.001$).

Overall, support messages for the MD & TD's (see Table 2) scenarios were rated highest, followed by ED & PD, then IN, FR and IS (significant using pairwise comparisons at $p<0.05$). This suggests that some scenarios are more easily supported.

Using pairwise comparisons, categories of support were formed into 5 distinct groups (significant from each other at $p<0.05$; see Table 6). These results partially support our hypothesis (**H2**) that empathetic, person-centred messages would be rated highest; however, Empathy (EM) was rated lower than expected.

The interaction effect of presented scenario×category can be seen in Figure 3. These results support our hypothesis (**H1**) that different categories of support will be more suitable in different scenarios. SUP and APP messages were rated highly for most scenarios, while PRS was suitable for scenarios when something had been accomplished. In addition, DES was rated highly for the Physical demand Scenario and ENC for the Mental Demand Scenario.

6.6.3. Effects of Scenario × Original Scenario.

We also investigated whether the messages were rated higher if the message being assessed was presented with the scenario it was created for in experiment 2. A 7×7 2-way ANOVA was performed on the effect of original scenario and presented scenario on rating. There was significant effect for original scenario ($F(6,11872)=113.90$, $p<0.01$), presented scenario ($F(6,11872)=17.63$, $p<0.01$) and for the interaction effect ($F(36,11872)=11.40$, $p<0.01$).

Using pairwise comparisons, we found that support messages originally generated from the PD & ED scenarios were rated highest, followed by TD; IS & MD; IN and FR (homogenous subsets significant from each other at $p<0.05$). Presented Scenario effects were as reported for Category × Scenario. The interaction effects can be seen in Figure 4. For the MD, TD & IS scenarios: messages provided for PD & ED were rated best. For PD: PD messages were rated best; for FR: ED messages were best; for IN: IN, PD and ED messages were best and for ED: PD and IS messages were rated highest (in all cases these were significantly better than other categories at $p<0.05$). These findings do not support our hypothesis (**H3**) that messages would be rated highest when they were presented with the scenarios for which they were produced, except for PD and IN. This suggests that for some stressful situations, people do not provide the most effective type of emotional support.

6.7. Algorithm Creation

Similar to how homogenous subsets of categories were created overall (as shown in Table 6 and explained in Section 6.6.2), we also created homoge-

[1]It is possible that there may be an impact from the variability in the number of ratings per participant, this is a limitation of this study.

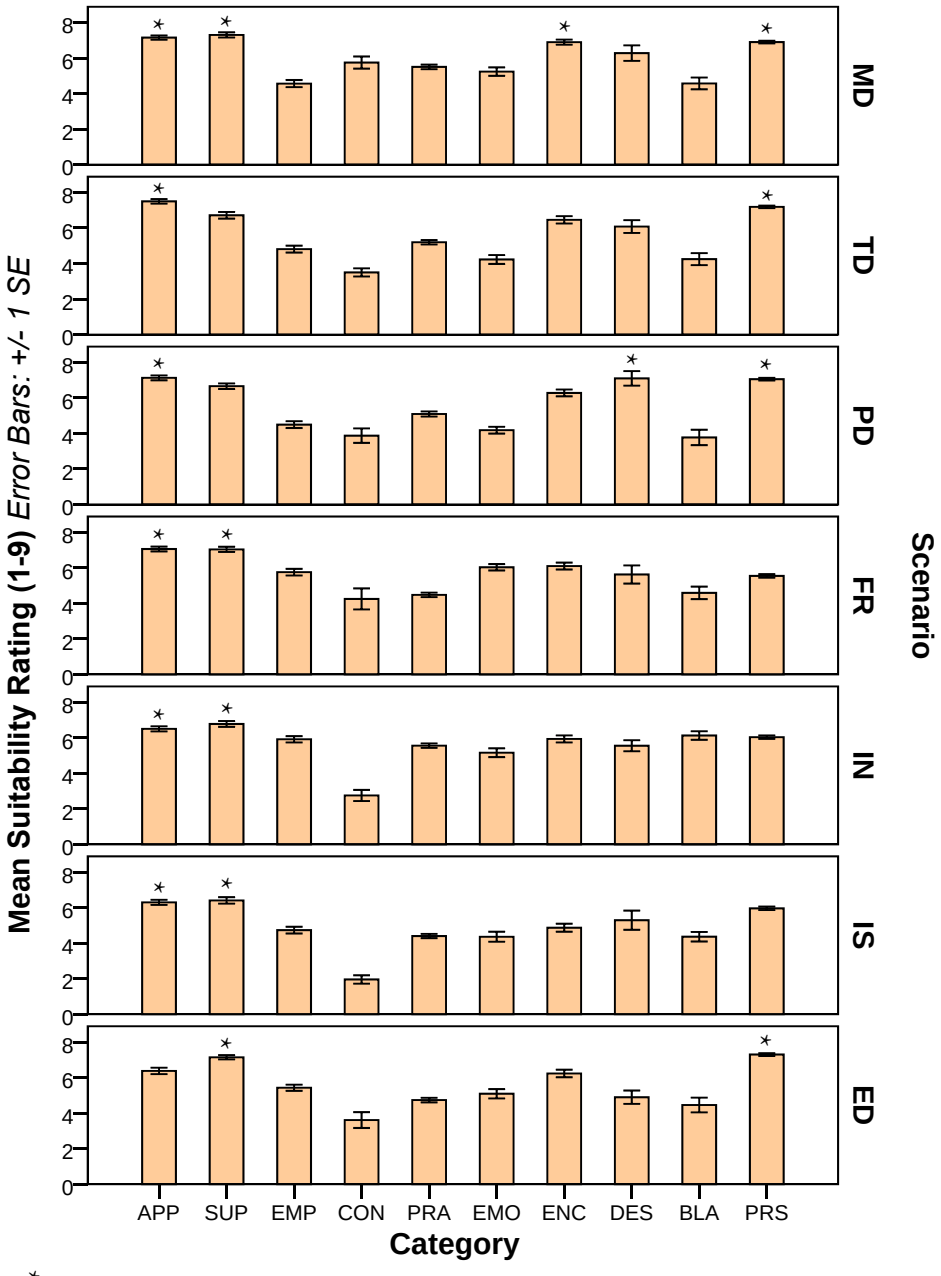

Fig. 3. Mean message suitability rating for each message category per scenario

Table 6
Homogenous Subsets of support from best to worst. Significant from each other using pairwise comparisons p<0.05

	Categories	Mean Suitability (1-9) & SE
1	Appreciated* (APP)	6.86 SE 0.06
	Supported* (SUP)	6.86 SE 0.06
	Praise* (PRS)	6.57 SE 0.03
2	Praise* (PRS)	6.57 SE 0.03
	Encouragement* (ENC)	6.11 SE 0.08
3	Encouragement* (ENC)	6.11 SE 0.08
	Deserving* (DES)	5.83 SE 0.16
4	Empathy* (EMP)	5.01 SE 0.06
	Practical Advice (PRA)	4.99 SE 0.05
	Emotional Advice (EMO)	4.90 SE 0.09
	Blameless (BLA)	4.59 SE 0.12
5	Consolation (CON)	3.67 SE 0.18

*Person-centred support categories [8]

neous subsets of categories (using pair-wise comparison) for each of the scenarios individually. For each scenario, the homogenous subset with the highest mean is indicated with stars in Figure 3.

Using these highest homogenous subsets per scenario, we created an algorithm of which type of support to provide to a carer experiencing different stressors (see Algorithm 1). For example, as can be seen in Figure 3, for Temporal Demand (TD), the support categories Appreciated (APP) and Praise (PRS) performed best. Therefore, for Temporal Demand the algorithm selects one out of these two categories randomly, and then randomly selects a support message from the selected category.

When the stressor is not known, the algorithm uses the highest homogeneous subset overall (Table 6) to select the categories.

7. Message Selection and Algorithm Refinement

The aim of this phase was to refine our support message set. For each stressor, the algorithm prescribes which support categories the message should come from. We therefore needed the best messages from each support category. As some support categories were used in the algorithm for multiple scenarios, we needed to ensure that the messages we selected would be suitable for all the scenarios in which they would be used.

Algorithm 1. Proposed initial algorithm

```
switch stressor_type do
    case MentalDemand:
        cats := {Supported, Appreciated,
                Praise, Encouragement}
    case TemporalDemand:
        cats := {Appreciated, Praise}
    case PhysicalDemand:
        cats := {Appreciated, Praise, Deserving}
    case EmotionalDemand:
        cats := {Supported, Praise}
    case Frustration:
    case Interruption:
    case Isolation:
        cats := {Supported, Appreciated}
    case Unknown:
        cats := {Supported, Appreciated, Praise}
category_to_show := select_random(cats)
messages := category_to_show->messages
message_to_show := select_random(messages)
show(message_to_show)
```

We examined the mean suitability rating for each support message per scenario. For each scenario, we selected a 'Whitelist' of all messages which had a mean Suitability rating (from Experiment 4) of at least 7.5 (these messages were 'highly Suited' for that scenario - indicated by '+' in Table 7). Additionally for

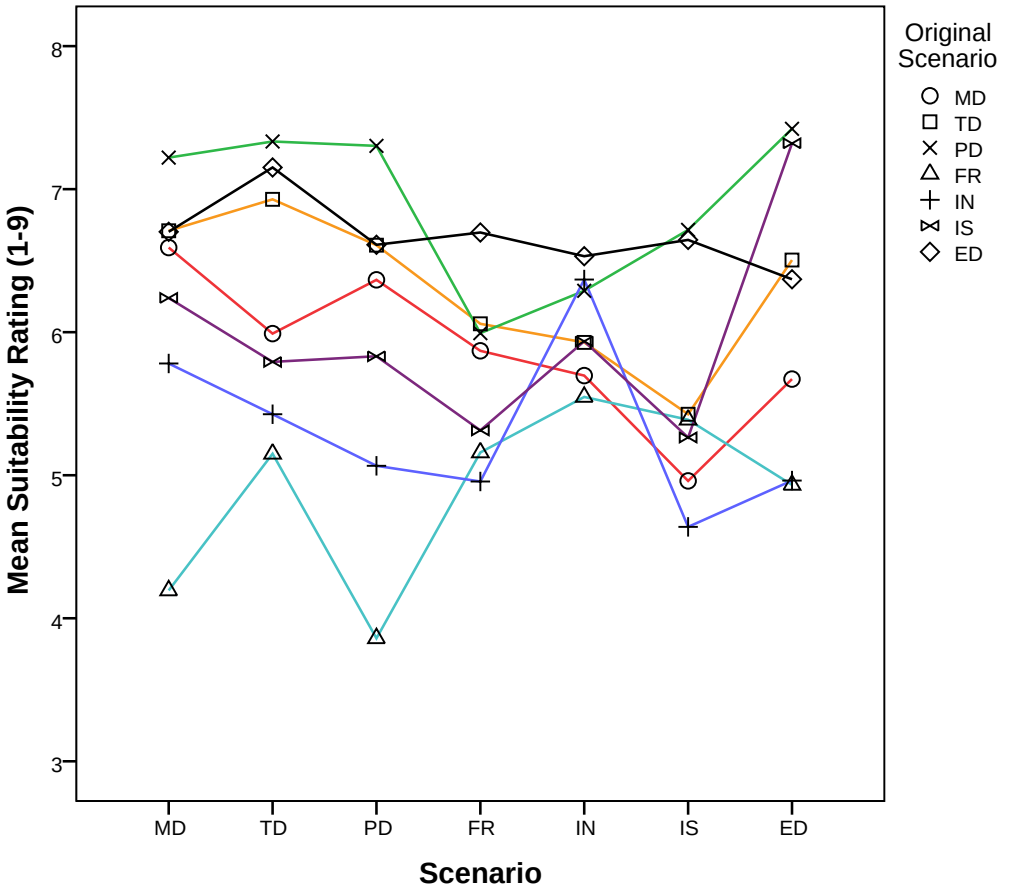

Fig. 4. Mean Message Rating per Presented Scenario against the Original Scenario they were generated for. Lines added for clarity only and do not indicate direction.

each scenario we compiled a 'Blacklist' of all messages which had a mean suitability rating of less than 5 (messages which were 'Poorly Suited' to that scenario; indicated by '-' in Table 7).

It transpired that some messages which were whitelisted for some scenarios were blacklisted for others. We overall blacklisted any messages that were blacklisted for any scenario that uses the message's category in the algorithm (Algorithm 1). Thus we only excluded blacklisted messages when they were blacklisted for a scenario to which that category applies e.g. 'Praise' message 'You were really kind today' is Blacklisted for the Isolation scenario, but was not excluded as 'Praise' messages are not recommended for the Isolation Scenario. This resulted in 2 messages being excluded.

To pick a set of the best messages for our algorithm, we compared the whitelists for each scenario. We wanted to ensure that the messages we were using were generalizable and not only good for one or 2 specific scenarios. Where a message appeared for at least half (rounded down) of the scenarios it should be applied to, it was added to our 'Best Messages' set (see Table 7). This resulted in a set of 3 'Appreciated' messages, 4 'Supported' messages, 0 'Encouragement' messages, 0 'Deserving' messages and 10 'Praise' messages (see Table 7). As 0 Encouragement or Deserving messages appeared on our 'Best Messages' list, the algorithm was refined to exclude these categories (see Algorithm 2).

8. Experiment Five: Algorithm Evaluation

Now that we have refined our algorithm and selected the best messages for it, we can evaluate it. This section describes two experiments which compared the

Table 7

All Messages with overall means and Blacklist(+) / Whitelist(-) per scenario. Greyed areas indicate that these categories are not recommended for this scenario by the algorithm

			Scenarios (+Whitelist, -Blacklist)									
Category	**Total Mean Suitability**	**Std. Error**	**MD**	**TD**	**PD**	**FR**	**IN**	**IS**	**ED**	**Blacklisted?**	**Best Message?**	**Message**
CON	3.67	0.163		-	-	-	-	-	-			At least that is over with, for the time being.
EMO	5.41	0.151			-			-				Just be calm
	5.25	0.155		-	-				-			Don't stress
	4.13	0.142	-	-	-		-	-	-			Don't get worked up about it.
ENC	6.57	0.133										You can do this.
	6.26	0.142										You will do great
	6.20	0.148		+								I know you will succeed
	5.32	0.141		-				-				You can find a way to deal with this
DES	5.83	0.172							-			You deserve some time off
BLA	5.02	0.162	-	-		-		-				It's not your fault
	3.99	0.166	-	-	-			-	-			There is nothing you can do about it
PRS	7.24	0.142		+	+				+		Y	You are invaluable
	7.17	0.141				+			+			Your dedication to Marker* is fantastic
	7.03	0.142	+		+						Y	You handled things well today.
	6.88	0.139		+	+				+		Y	You did a good thing
	6.82	0.130	+		+						Y	You are a wonderful carer.
	6.80	0.138			+			-	+		Y	You were really kind today
	6.78	0.130	+		+						Y	Your effort is commendable
	6.69	0.144		+	+						Y	You are a wonder
	6.68	0.145	+	+							Y	You did an excellent job today
	6.66	0.138							+			You're very good at your job.
	6.66	0.139		+		-			+		Y	You are an amazing person
	6.52	0.142					+					You're very patient.
	6.47	0.134										You are a good person.
	6.45	0.136							+			You are awesome.
	6.41	0.138		+	+			-			Y	You are a hard worker
	6.34	0.138				-			+			You are a hero.
	6.11	0.141				-						You are really good at managing your time.
	6.04	0.141				-						You are a strong person
	5.92	0.141				-						You are a special person
	5.91	0.138							+			Your understanding is admirable
APP	7.70	0.148		+	+		+		+		Y	Marker's* really lucky to have you.
	7.27	0.137		+				+	+			Marker* appreciates you.
	7.11	0.142	+	+	+	+					Y	I'm really glad you are here for Marker*.
	6.99	0.147		+	+	+					Y	Your work is very appreciated.
	6.67	0.147		+								I'm really glad you are here for Marker*.
	6.49	0.144	+	+					-			Marker's* life is better for you
	6.45	0.141	+	+				-		Y		Marker* couldn't make do without you.
	5.93	0.147										Marker* is grateful to you.
SUP	7.33	0.137	+		+				+		Y	Call me whenever you feel overwhelmed.
	7.19	0.138	+				+				Y	I am available if you need assistance.
	7.17	0.134		+		+	+				Y	I am here for you.
	7.03	0.135	+		+	+	+				Y	Let me help you
	6.84	0.127						+				Please feel free to call and talk to me anytime.
	5.02	0.147		-	-	-	-		+	Y		Tell me about the things you can't say to Marker*, that you keep to yourself.
EMP	6.65	0.155			+							I understand how stressful it must be.
	5.69	0.154						-				Wow that must have been hard
	5.02	0.145		-	-	-	-	-				How are you doing after that?
	4.86	0.151	-		-		-					Oh I'm sorry to hear that
	4.44	0.160	-	-	-				-			I understand that must have been disappointing.
	4.42	0.154	-	-	-		+	-	-			That's really frustrating
PRA	6.54	0.153			+	-						Just take it one step at a time
	5.79	0.154						-				Do your best and prioritize.
	5.44	0.152						-				Breathe.
	5.43	0.157		-		-						Take it slow.
	4.97	0.169			-	-		-				Make a plan and make it happen.
	4.94	0.148			-	-		-	-			Maintain focus
	4.83	0.150				-	-	-	-			Be careful.
	4.71	0.150	-	-	-	-		-	-			Ignore those things that can wait.
	4.57	0.154	-	-	-	-		-	-			Focus on priorities.
	4.30	0.155		-		-	-	-	-			Practice makes perfect.
	4.21	0.167	-	-	-		-	-	-			Just get through it.
	3.89	0.160	-	-	-	-		-				Try to concentrate.

Algorithm 2. Proposed refined algorithm

```
switch stressor_type do
    case MentalDemand:
        cats := {Supported, Appreciated, Praise}
    case TemporalDemand:
    case PhysicalDemand:
        cats := {Appreciated, Praise}
    case EmotionalDemand:
        cats := {Supported, Praise}
    case Frustration:
    case Interruption:
    case Isolation:
        cats := {Supported, Appreciated}
    case Unknown:
        cats := {Supported, Appreciated, Praise}
category_to_show := select_random(cats)
messages := category_to_show->messages
message_to_show := select_random(messages)
show(message_to_show)
```

Table 8
Messages selected for Experiment 5

Category	Message
Appreciated	Your work is very appreciated.
Supported	Let me help you
Empathy	I understand how stressful it must be
Practical Advice	Just take it one step at a time
Encouragement	You can do this.
Praise	You are an amazing person

predictions of our algorithm on which message to use for each scenario with the preferences of participants.

8.1. Experiment Five A

In the first experiment, we decided to allow the participants to choose from a set of emotional support messages which one was most suitable in a given scenario so that we could compare whether the prediction of our algorithm matched their message choice.

8.1.1. Design

We used a within-subject design: each participant considered all scenarios. The independent variable was Scenario (7 levels) and the dependent variable was Message (6 levels).

8.1.2. Hypotheses

- **H1**: Messages more frequently selected for each scenario will be the same as predicted by the algorithm (see Algorithm 2).

8.1.3. Participants

There were 30 participants: 14 female and 16 male. Four were aged 18-25, 20 were 26-40 and 6 aged 41-65.

8.1.4. Materials

We used the seven validated scenarios from Experiment One (see Table 2).

We used six messages (see Table 8). To select the messages to use, we first selected 1 message for each of the 10 categories from Table 7 (using a message from our 'Best Message' set whenever such a set existed i.e. for Praise, Appreciated and Supported). We attempted to select messages with a similar overall mean suitability.[2] To verify this, a one-way ANOVA was performed between the 10 messages on Suitability rating (from Experiment Four). The ANOVA was significant at $F(1827)=44.17$ $p<0.001$. Post-hoc test revealed that the messages selected for CON, EMO, DES and BLA had significantly lower overall ratings ($p<0.05$) than APP, SUP, EMP, PRA, ENC and PRS, which did not differ from each other. We thus excluded the CON, EMO, DES and BLA messages from our evaluation (these messages would have performed badly not due to the scenario but due to their overall quality), leaving us with 6 messages (see Table 8).

8.1.5. Procedure

Each participant was presented with each scenario in turn and a radio button to select the most suitable emotional support message (from the set of 6) that they would like to provide. A comments box was provided to explain why they had made that choice.

8.1.6. Results

A Chi Squared test was performed on Scenario$\times$ Message. This was significant at χ^2 $(30)=126.29$, $p<0.001$. Adjusted residuals were examined to see which messages were most frequently selected for each scenario. The results can be seen in Table 9 and the messages chosen significantly more frequently for each scenario are summarised in Table 10.

It was found that the Algorithm (Algorithm 1) alone did not predict the messages that would be selected most frequently for each scenario; thus we fail to find

[2]We wanted to ensure that the effect was due to the category and was not biased by the overall quality of the message.

Table 9
χ^2 of Message Choice counts per Scenario

	Message Category						
Scenario	APP	SUP	EMP	PRA	ENC	PRS	Total
MD	5	0*	1*	13*	11*	0*	30
TD	6	2	6	7	4	5	30
PD	8	10*	1*	2	2	7	30
FR	4	10*	10*	2	2	2	30
IN	2*	5	13*	4	6	0*	30
IS	15*	1	1*	1*	7	5	30
ED	6	1	6	4	2	11*	30
Total	46	29	38	33	34	30	210

*Adjusted residual $\geq \pm$ 2.0 ; This score is significantly different than predicted.

Table 10
Messages selected for each Scenario

Scenario	Best predicted categories from algorithm (removed after refinement)	Best predicted categories from individual messages	*Experiment 5A* Messages Selected (χ^2 adjusted residual$\geq\pm$2.0)	*Experiment 5B* Messages Ranked (Significant using Pairwise Comparisons p<0.05)
MD	APP, SUP, PRS, (ENC)	APP, SUP, PRA, PRS	PRA‡, ENC†	ENC†, PRA‡, APP†‡, EMP*
TD	APP, PRS	APP, PRS, PRA	*Messages NS different*	APP†, PRS†, SUP*, EMP*, ENC*
PD	APP, PRS, (DES)	APP, SUP, EMP, PRA	SUP‡	SUP‡
FR	APP, SUP	APP, SUP	SUP†, EMP*	SUP†‡, APP†‡ EMP*
IN	APP, SUP	SUP, EMP	EMP‡	SUP†‡, EMP‡
IS	APP, SUP	APP, SUP, EMP	APP†‡	APP†‡, EMP‡
ED	PRS, SUP	PRS	PRS†‡	SUP†, PRS†, EMP*, APP*

† Predicted by the algorithm

‡ Predicted by the individual sentence performance

*Not predicted by either the algorithm or the individual sentence performance

evidence for **H1**. To explain this, we re-examined the individual messages' performance for each scenario using our data from Experiment Four. We ran a 7×6 2-way ANOVA of Scenario×Message on Suitability rating. This was significant at F(30,1114)=6.77, p<0.001. Pairwise comparisons showed that the best of these messages for each scenario was slightly different than our algorithm predicted e.g. for the Interruption Scenario, the individual messages that were best were SUP and EMP while the algorithm recommends APP and SUP. These are shown in Table 10. This factored in, the Algorithm and Individual Message performance predict most of our results.

8.2. Experiment Five B

We modified Experiment Five A in three ways. Firstly, Experiment Five B focused on which support participants would like to *receive*, rather than which support they would like to give (as in Experiment Five A). The rationale for this was that there is likely to be a difference between when one is asked to provide support than when one is asked to reflect on the support one would like to receive (some people are not good at providing effective emotional support [17]). Support for this is provided by the results of Experiment Four, where sometimes messages originally produced from a scenario were rated lower for that scenario than messages originally produced for a different scenario.

Secondly, Experiment Five B allowed participants to rank all six messages rather than forcing the selection of just one message (as in Experiment Five A). The rationale was that several messages may be almost as preferable, and it would help to know how well the messages recommended by the algorithm did compared to all messages in the set (in Experiment Five A, we would not be able to distinguish whether the message chosen by the algorithm was second best or worst for a given participant).

Thirdly, Experiment Five B asked participants to rate how well they empathised with the scenario. The rationale for this is that we hypothesized that the amount an individual empathises with a situation will affect how highly they rate the message - Jones and Burleson [14] found that people viewed low-person-centered messages as more appropriate when the message recipient was viewed as blameworthy and thus empathized with less. Thus there might be an effect of empathy on people's choice of message categories.

8.2.1. Design

We used a within-subject design: each participant considered all scenarios. The independent variables were Scenario (7 levels) and Message (6).

The dependent variables were:

- Message rank: This indicated the participant's relative preference between the support messages (from options '1 First', '2 Second', etc.). No two messages could be given an equal rank. They could also choose not to rank any message as 'I wouldn't like this support' (7). Before analysis, message rank was recoded so that '1 First' was recoded as 6, '2 Second' as 5... '6 Sixth' as 1 and 'I wouldn't like this support' as 0.
- Sympathy: This measures how well participants thought they could empathise with the stress the carer had experienced on a scale of 1-7 (1 Very poorly = 'I don't understand this situation/would not find this stressful' & 7 Very well='I have experienced a similar situation and understand exactly how stressful it is'). To disambiguate it from the Message Category 'Empathy', this variable was named 'Sympathy'. For the analysis, Sympathy was divided into 3 groups - Low (ratings 1-2), Medium (ratings 3-5) and High (ratings 6-7).

8.2.2. Hypotheses

- **H1**: Messages ranked more highly for each scenario will be predicted by the algorithm (as before)
- **H2**: Participants with higher sympathy (Medium or High) will rank person-centred messages (Appreciated, Praise, Supported, Empathy, Encouragement) higher than participants with low sympathy.

8.2.3. Participants

There were 31 participants: 18 female, 13 male. Four were aged 18-25, 21 were 26-40 and 6 were 41-65.

8.2.4. Materials

The same 7 scenarios and 6 messages were used as in Experiment Five A.

8.2.5. Procedure

Each participant was presented with each scenario in turn. First, they were asked to rate how well they thought they could empathise with the stress the carer had experienced on a scale of 1-7 (Sympathy). Next, they were asked to imagine they were the carer and rank the support messages they would like to receive (they could also choose not to rank any message as 'I wouldn't like this support'). A comments box was provided to explain why they had given those rankings.

8.2.6. Results

Effects of Scenario×Sympathy From Experiment Four we found that support messages for the MD & TD's scenarios were rated highest, followed by ED & PD, then IN, FR and IS. Thus we were interested in seeing whether the scenarios whose messages were rated lowest were the scenarios which people had the lowest Sympathy for. A 1-way ANOVA was performed of Scenario on Sympathy rating. This was significant at $F(6,203)=3.24$, $p=0.005$. Using pairwise comparisons, the lowest homogenous subset of Scenarios was IS, FR, IN and ED (see Figure 5). This is precisely what we expected, based on Experiment 4.

This result implies two things: Firstly, people do provide worse support messages for scenarios that they do not empathise with and secondly that the stressors people find hardest to empathise with are FR, IN, IS and ED. These scenarios are therefore potentially the most important for our algorithm to perform well at, as they are likely to be the least well supported by friends and family.

Effects of Scenario×Message We were interested in seeing what people who could envisage themselves in the carer's position would rate as good support. The low sympathy group was thus excluded from analysis of Scenario×Message. A 2-way 7×6 ANOVA was performed on Scenario×Message. There was a main effect of Scenario×Message at $F(30, 1086)=3.08$, $p<0.001$. Pairwise comparisons revealed the best ranked Messages for each Scenario (see Figure 6 and Table 10). Overall the messages ranked most highly are predicted either by individual message performance (from Experiment Four) or the algorithm. The main exception to this is the empathy statement, which performs well across all the scenarios.

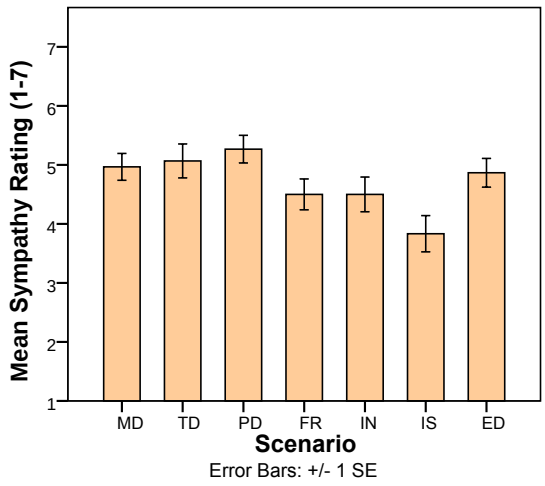

Fig. 5. Effect of Scenario on Sympathy Rating

This result provides partial support for **H1**: the algorithm provides good predictions but a final system will need to consider the performance of individual messages as well.

Effects of Message×Sympathy In order to investigate if message choice varies with sympathy, a 6×2 2-way ANOVA was performed on Message×Sympathy. This was significant at F(10,1134)=1.84, p<0.05. There were also individual effects for Message (F(5,1176)= 3.70, p<0.005) and Sympathy (F(2,1176)=5.52, $p < 0.005$).

Multiple comparisons of Message show that the Appreciated (APP) and Supported (SUP) messages are ranked the highest (significant at p<0.05). This is consistent with our results from Experiment Four. Additionally we found that Medium and High Sympathy groups have significantly higher average rank score than the low sympathy group (p<0.05). This implies that the low sympathy group ranked more messages as 'I wouldn't like this support', ranking fewer messages overall.

Pairwise comparisons reveal that Medium and High sympathy groups ranked Appreciated (APP), Empathy (EMP) and Praise (PRS) messages significantly more highly than the low sympathy group (p<0.05). We also found that in low sympathy groups the only significant difference in ranking of message types was that Supported (SUP) was ranked higher than Praise (PRS), while for Medium and high sympathy groups, Appreciated (APP) and Supported (SUP) were ranked higher than all other message categories (see Figure 7).

The results provide partial support for **H2**: for three of the person-centered messages (APP, EMP, PRS), there was indeed a significant difference for the degree of empathy. The results also suggest that increased empathy not only increases the *amount* of support wanted (with fewer 'I wouldn't like this support' selections), but also specialises the *types* of support preferred.

9. Discussion

Our results fit well with [4]'s model (See Figure 1) of supportive behaviours where support is either approach/avoid and emotion/problem-focused. If we divide our 10 categories (see Table 5) into these strategies by comparing the definitions of our categories to [4]'s (as discussed in Section 1), *Appreciated*, *Supported* and *Empathy* fall under 'Solace'; *Praise*, *Encouragement* and *Deserving* are 'Solace'/'Solve' (they elicit positive emotion whilst encouraging a solution); *Practical Advice* and *Emotional Advice* are 'Solve'; *Blameless* is 'Escape' and Consolation is 'Dismiss'. We found that the categories under 'Solace' performed best, followed by 'Solve', 'Escape' and 'Dismiss'. The only inconsistent category was *Empathy*, which was not highly rated in Experiment Four. It was however, highly rated in Experiment Five B. It is probable that as empathy requires the supporter to acknowledge and express how the individual is feeling, many empathy messages must be scenario-specific and thus performed badly when matched with other scenarios (we removed scenario-specific messages in creating our message corpus, so we may have removed high quality empathic support). In Experiment Five, we selected an empathy message to test that was generalizable and thus showed that Empathy is an effective support message category.

Burleson's [8] framework of person-centredness is also supported by our results - categories which describe, legitimise and sympathise with the distress generally performed better than categories which advise how to act or feel (PRA, EMO), or which distract them from the distress (BLA, CON). This framework does not however encapsulate the success of simply offering support or telling someone they are appreciated. Within the field of carer support (where isolation from friends and degradation of social ties is a key problem), reminding someone that their friends are there for them and they are appreciated by their caree should be a principal feature of their emotional support.

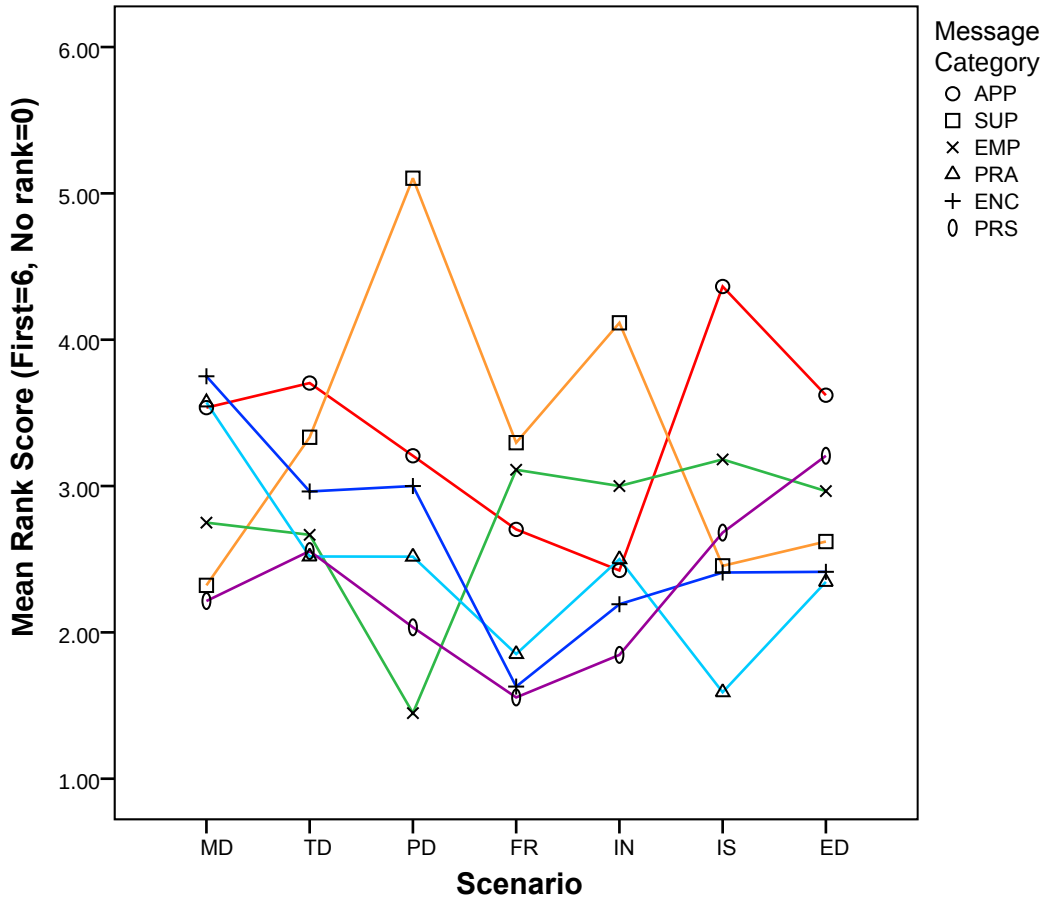

Fig. 6. Mean Message Rank per Scenario for each Message. Lines added for clarity only and do not indicate direction.

We found that the best support messages for a scenario were not the ones originally intended for that scenario, except for the Physical Demand and Interruption Scenarios. It is possible that people produce the best support for situations they empathize with most - Jones and Burleson [14] found that people viewed low-person-centered messages as more appropriate when the message recipient was viewed as blameworthy and thus empathized with less. As such, high empathy scenarios should be best for eliciting high quality general emotional support messages, while for Physical Demand and Interruption scenarios, scenario-specific support messages should be used. It is possible that in certain scenarios, a virtual agent will outperform a human for simple emotional support.

Our results also show that *Praise* is an effective support type for Mental, Temporal, Physical and Emotional Demand. Reflection on our scenario content (see Table 2) suggests that this may be because some activity was achieved in these scenarios; further investigation is needed to determine whether achievement is also a factor when selecting emotional support. It would also be useful to explore why *Deserving* messages (e.g. "You deserve some time off") were suitable for the Physical Demand scenario and why *Encouragement* messages (e.g. "You will do great") were suitable for the Mental Demand scenario.

Our evaluation suggests that there is an impact of the content of the specific message on message suitability. This is expected - messages are likely to have slightly different connotations not reflected by our coarse message categories. Although our predictions were not fully reflected by our results, this does not mean that our algorithm performs poorly - rather that certain specific messages in specific situations are better suited than our algorithm's recommendations. However, we do not anticipate that our Agent will be able to analyse

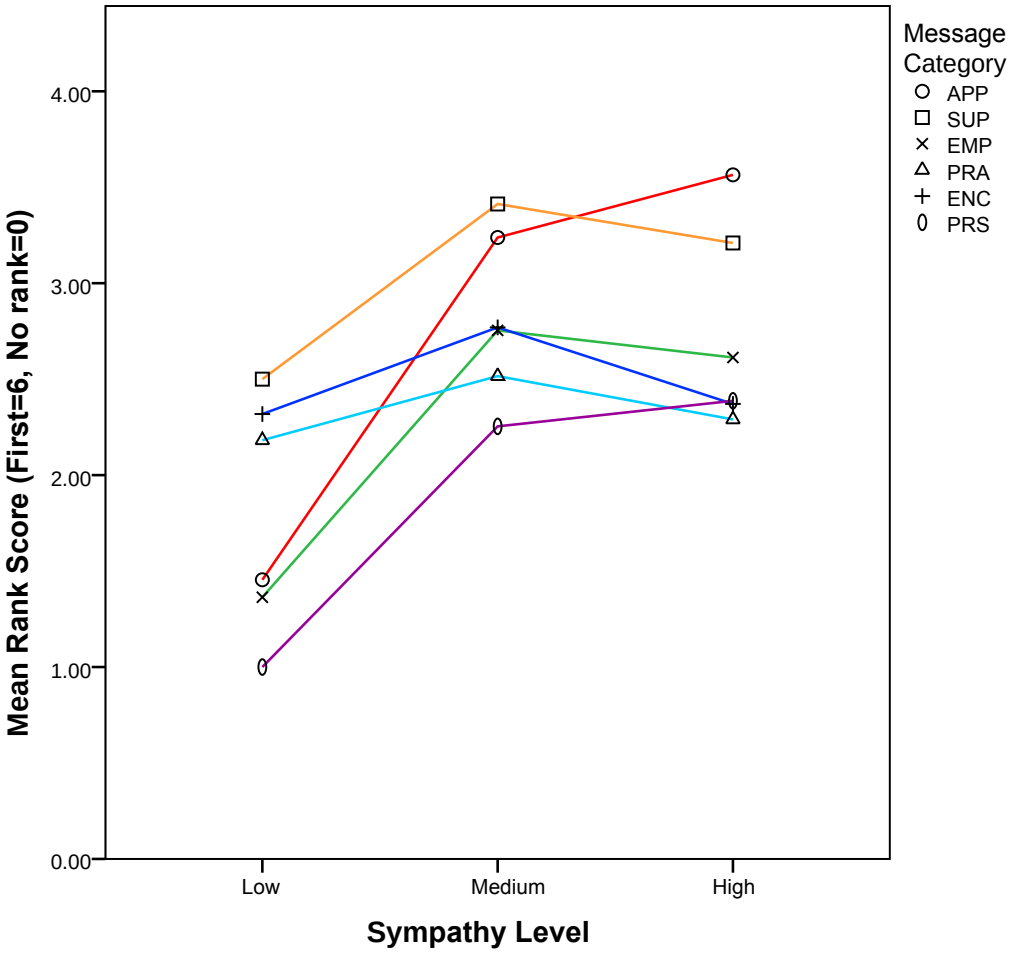

Fig. 7. Effect of Message Category × Sympathy on Mean Message Ranking

the carer's situation in enough detail to make such a fine-grained approach plausible.

We did not investigate the impact of gender in this paper, but it may have had an impact. Different people have different support seeking behaviour e.g. males are less likely to seek emotional support than instrumental support [3]. Thus participants may have rated instrumental support higher for a male carer. Furthermore, people differ in how good they are at providing high quality emotional support e.g. females are generally better supporters [17] - using participants who were good at providing emotional support may have provided us with a higher quality message set. Cultural differences may also have an impact.

Personality has an impact on perceived supportedness [2] and support reciprocity [16]; therefore it may have an impact on the most suitable support, as it does for learners [11]. Future work will therefore also investigate how to adapt messages to carer personality. Additionally, combining message categories might improve support by offering both solace and solve support.

From our results, we have developed, refined and evaluated an algorithm for providing emotional support for carers in stressful situations (see Algorithm 2). Future work will involve creating more scenarios depicting our stressors so we can be sure that it is the stressor and not other features of the scenario content that make the messages suitable. We also intend to use the best rated messages to generate a larger corpus of support messages as some of our categories contained very few messages (see Table 5) and this may have influenced how the category was rated. This will allow us to create a virtual agent that can produce suitable,

high quality emotional support messages for a carer, adaptive to their situation.

Currently, the algorithm does not really provide for multi-stressor scenarios, though a recommendation has been added for when the stressor is not known. We need to explore the impact of combinations of specific stressors to extend our algorithm.

Once this algorithm has been implemented with a virtual agent in carer support environment, we intend to test it with carers experiencing stress.

10. Conclusion

In this paper we investigated which types of emotional support messages are most suitable to offer carers affected by different types of stressors. We were then able to create an algorithm that a virtual agent could use to offer emotional support. We collected a corpus of 63 emotional support messages categorized into 10 categories and 7 care scenarios depicting 7 different stressors. We found that support was rated differently across scenarios, and that there was an interaction between scenario and support category. Empathetic, person-centred messages were rated highly, especially *Supported* and *Appreciated*. However, the specific category *Empathy* was rated lower than expected. However, exploring this issue revealed that one Empathy statement was consistently highly ranked. Potentially other Empathy messages were rated low because empathy is intrinsically scenario-specific.

Surprisingly, messages were not always considered most suitable when they were presented with the scenarios for which they were produced, with the exception of *Physical Demand* and *Interruption*. This suggests that for some stressful situations, people do not provide the most effective type of emotional support. Exploring this issue, we found that how well people feel they can understand the situation affects the type and quantity of support they give.

The evaluation of our algorithm suggests that there is an impact of the content of the specific message on message suitability. This is expected - messages are likely to have slightly different connotations not reflected by our coarse message categories. Although our predictions were not fully reflected by our results, our algorithm still performs well.

These results suggest that there is a promising scope for IVA's to tailor emotional support to a carer's personal situation, and may in some cases perform better than humans.

11. Acknowledgments

This paper acknowledges the Northern Research Partnership.

References

[1] Census 2001. Office for National Statistics (2001)
[2] Asendorpf, J.B., Wilpers, S.: Personality effects on social relationships. Journal of Personality and Social Psychology 74(6), 1531 (1998)
[3] Ashton, W.A., Fuehrer, A.: Effects of gender and gender role identification of participant and type of social support resource on support seeking. Sex roles 28(7-8), 461–476 (1993)
[4] Barbee, A.P., Cunningham, M.R., Winstead, B.A., Derlega, V.J., Gulley, M.R., Yankeelov, P.A., Druen, P.B.: Effects of gender role expectations on the social support process. Journal of Social Issues 49(3), 175–190 (1993)
[5] Barbee, A., Cunningham, M.: An experimental approach to social support communications: Interactive coping in close relationships. In: Communication yearbook. vol. 18, pp. 381–413 (1995)
[6] Bloom, J.R., Spiegel, D.: The relationship of two dimensions of social support to the psychological well-being and social functioning of women with advanced breast cancer. Social Science & Medicine 19(8), 831–837 (Jan 1984)
[7] Buckner, L. & Yeandle, S.: Valuing carers 2011. Carers UK, London (2011)
[8] Burleson, B.R., Daly, J., Wiemann, J.: Comforting messages: Features, functions, and outcomes. Strategic interpersonal communication pp. 135–161 (1994)
[9] Dennis, M., Kindness, P., Masthoff, J., Mellish, C., Smith, K.: Towards effective emotional support for community first responders experiencing stress. Humaine Association Conference on Affective Computing and Intelligent Interaction (2013)
[10] Dennis, M., Masthoff, J., Mellish, C.: The quest for validated personality trait stories. In: IUI. pp. 273–276. ACM (2012)
[11] Dennis, M., Masthoff, J., Mellish, C.: Does learner conscientiousness matter when generating emotional support in feedback? Humaine Association Conference on Affective Computing and Intelligent Interaction (2013)
[12] Gross, J.J.: The emerging field of emotion regulation: An integrative review. Review of general psychology 2(3), 271 (1998)
[13] Hart, S.G.: Nasa-task load index (nasa-tlx); 20 years later. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. vol. 50, pp. 904–908. Sage Publications (2006)
[14] Jones, S.M., Bureleson, B.R.: The impact of situational variables on helpers' perceptions of comforting messages: An attributional analysis. Communication Research 24(5), 530–555 (Oct 1997)
[15] Klein, J., Moon, Y., Picard, R.W.: This computer responds to user frustration:: Theory, design, and results. Interacting with computers 14(2), 119–140 (2002)
[16] Knoll, N., Burkert, S., Schwarzer, R.: Reciprocal support provision: Personality as a moderator? European journal of personality 20(3), 217–236 (2006)
[17] MacGeorge, E.L., Gillihan, S.J., Samter, W., Clark, R.A.: Skill deficit or differential motivation? testing alternative explana-

tions for gender differences in the provision of emotional support. Communication Research 30(3), 272–303 (2003)
[18] Masthoff, J.: The user as wizard: A method for early involvement in the design and evaluation of adaptive systems. Workshop on User-Centred Design and Evaluation of Adaptive Systems pp. 460–469 (2006)
[19] Meyer, D.K., Turner, J.C.: Discovering emotion in classroom motivation research. Educational psychologist 37(2), 107–114 (2002)
[20] MT: Amazon mechanical turk. http://www.mturk.com
[21] Prendinger, H., Ishizuka, M.: The empathic companion: A character-based interface that addresses users'affective states. Applied Artificial Intelligence 19(3-4), 267–285 (2005)
[22] Randolph, J.J.: Free-marginal multirater kappa: An alternative to fleiss' fixed-marginal multirater kappa. In: Joensuu University Learning and Instruction Symposium 2005 (2005)
[23] de Rosis, F., Novielli, N., Carofiglio, V., Cavalluzzi, A., De Carolis, B.: User modeling and adaptation in health promotion dialogs with an animated character. J Biomed Inform 39(5), 514–531 (2006)
[24] Taylor, W.L.: Cloze procedure: A new tool for measuring readability. Journalism Quarterly 30, 415–433 (1953)
[25] Vitaliano, P.P., Zhang, J., Scanlan, J.M.: Is caregiving hazardous to one's physical health? a meta-analysis. Psychological Bulletin 129(6), 946–72 (2003)
[26] Wang, C.Y., Chen, G.D., Liu, C.C., Liu, B.J.: Design an empathic virtual human to encourage and persuade learners in e-learning systems. In: Proceedings of the first ACM international workshop on Multimedia technologies for distance learning. pp. 27–32. ACM (2009)