Robot vs. Human Doctor (Technical Report)

Sonja D. Winter and Johan F. Hoorn

Vrije Universiteit Amsterdam

Abstract

Obviously, patients like it best when things "go well" (positive wording, positive outcome) but they dislike most a positive outcome that is brought negatively ("The news is not bad"). If things go badly, they do not want to get it straight into their face ("It's going bad") but rather prefer the denial of the positive outcome ("It's not going too well"). If done the wrong way, people will feel negative about the doctor, about the health message, expect a lower quality of life, and do not follow up on doctor's advice (Burgers, Beukeboom, & Sparks, 2012). Repeating this study with the positive-negative framing and affirmative-negative language with 134 participants receiving bad news from a robot doctor about Bekhterev's disease, we expected effects to be quite similar, albeit on a lower level. But none of our hypotheses were corroborated. For human doctors, language and framing make a difference. But robot doctors can say it any way they want. They performed better than human doctors throughout at *Doctor Evaluation, Message Evaluation, Expected Quality of Life,* and *Medical Adherence*. Additionally, we measured how the robot was experienced in terms of *Ethics, Affordances, Involvement, Distance,* and *Use Intentions.* Doctor Robot was not seen as distant but rather involving, morally good, skillful, and evoking the willingness to consult her again. The current analyses show how we came to these unexpected results.

Table of Contents

Scale Analysis

Message-related outcome variables

To assess the factor structure of our four outcome variables *Doctor Evaluation, Message Evaluation, Expected Quality of Life,* and *Medical Adherence*, we executed EFA (Maximum likelihood estimation) with Promax rotation, expecting 4 factors. Bartlett's Test of Sphericity was significant (χ^2 = 836.21, p < .001) and the Kaiser-Meyer-Olkin Measure of Sampling Adequacy adequately high at .775 (should be > .600). Both these measures indicated that executing an EFA would be reasonable for these data. Model fit was good (χ^2= 85.75, p <.001, RMSEA=.066). The pattern matrix of Table 1 shows the loadings of each item onto each factor. For clarity, loadings < .3 were removed.

Table 1. EFA factor loadings with Promax rotation: outcome variables

Pattern Matrix[a]

Item	Label	Factor			
		1	2	3	4
Cmnalf1	Ik ga proberen de instructies van de dokter op te volgen.			0.637	
Cmnalf2	Het opvolgen van de adviezen van de dokter is verstandig.			1.049	
Cmnalf3	Het is een goed idee om de adviezen voor behandeling op te volgen.			0.767	
ctkmstdr	Recoded: Ik denk dat de dokter mijn levenskwaliteit laag inschat.				0.821
ctkmstzl	Op basis van dit gesprek schat ik mijn levenskwaliteit hoog in.				-0.669
Cib1	Was begrijpelijk		0.865		
Cib2	Was duidelijk		0.917		
Cib3	Recoded: Nam mijn hoop weg				0.408
Cib4	Was informatief		0.670		
Cid1	Beleefd	0.875			
Cid2	Meelevend	0.600			

Cid3	Respectvol	0.760			
Cid4	Tactvol	0.631			
Cid5	Vriendelijk	0.730			

Extraction Method: Maximum Likelihood.

Rotation Method: Promax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

All items neatly loaded onto their own scale, with the exception of the recoded *Message Evaluation* item "Nam mijn hoop weg" ("Took my hope away") and the *Expected Quality of Life* item "Op basis van dit gesprek schat ik mijn levenskwaliteit hoog in" ("On the basis of this conversation I expect the quality of my life will be high"). The 'hope' item loaded quite high on the same factor as the second *Expected Quality of Life* item. Indeed, taking someone's hope away also has a link with future feelings and quality of life.

The 'high quality of life' item had a surprisingly negative loading on the *Expected Quality of Life* scale. This is strange, because this item was worded contrary to the second *Expected Quality of Life* item. We recoded the second item ("ctkmstdr") prior to EFA, expecting that this would result in a positive correlation between these two items (as opposed to a negative, we expected one positive and one negative item). However, the results showed that the recoded variable actually had a negative association with the other item. A separate correlation analysis confirmed that the original items (before recoding) had a positive correlation ($r = .572$, $p < .001$). This would indicate that as participants' evaluated the doctor's take on their quality of life as more positive, they evaluated their own view on their quality of life as less positive.

Based on these two findings, it seemed the best solution to combine the two recoded items ("ctkmstdr" and "cib3") in a new *Expected Quality of Life* scale, and to drop the other *Expected Quality of Life* item ("ctkmstzl").

This left us with the following scales: *Message Evaluation* (3 items, Cronbach's alpha = .889), *Doctor Evaluation* (5 items, Cronbach's alpha = .884), *Expected Quality of Life* (2 items, one originally from *Message Evaluation*, Cronbach's alpha = .653), *Medical Adherence* (2 items; Cronbach's alpha = .842).

The use of the Promax rotation was justified by the correlation matrix of the

solution (Table 2).

Table 2. Correlation Matrix of EFA

Factor Correlation Matrix

Factor	1	2	3	4
1	1.000	.462	.469	-.097
2	.462	1.000	.514	-.077
3	.469	.514	1.000	-.106
4	-.097	-.077	-.106	1.000

Extraction Method: Maximum Likelihood.

Rotation Method: Promax with Kaiser Normalization.

Ethics, Involvement, and Distance in the Robot Doctor condition (N=134)

The experiential variables in the Robot Doctor study (*n*=134) were *Ethics, Involvement,* and *Distance*. To assess the factor structure of these three variables, we executed EFA (Maximum likelihood estimation) with Promax rotation, expecting 3 factors. Bartlett's Test of Sphericity was significant (χ^2 = 1045.08, $p < .001$) and the Kaiser-Meyer-Olkin Measure of Sampling Adequacy adequately high at .880 (should be > .600). Both these measures indicated that executing an EFA made sense for these data. Model fit was good (χ^2= 102.15, p =.001, RMSEA=.068).

The pattern matrix below (Table 3) shows the loadings of each item onto each factor. Loadings < .3 were removed for clarity.

Table 3. EFA factor loadings with Promax rotation: *Ethics, Involvement, Distance*

Pattern Matrix[a]

	Factor		
	1	2	3
DistA			.652
DistB			.386
DistC			.758
DistD			.883
InvA		.535	

InvB		.535	
InvC		.579	-.350
InvD		.500	
EthA	.819		
EthB		.830	
EthC	.749		
EthD	.887		
EthE	.773		
EthF		.605	
EthG	.639		

Extraction Method: Maximum Likelihood.

Rotation Method: Promax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

The general factor structure looked good, with some exceptions. First, EthB ("Deze robot-dokter is aardig" "This robot doctor is nice") and EthF ("Deze robot-dokter is goedwillend" "This robot doctor is benevolent") had a low loading on their own *Ethics* factor, but a high loading on the *Involvement* factor. Looking at the content of these items, it is not unexpected that they cling to the *Involvement* items (e.g. "Ik heb een goed gevoel bij deze dokter" "I have a good feeling about this doctor"), thus we added them to the involvement scale.

Furthermore, InvC had a relatively low, negative cross loading on the *Distance* factor. As this cross-loading was relatively low, and its loading on the *Involvement* scale was higher and positive, we decided to keep this item as part of the *Involvement* scale.

This left us with the following scales: *Ethics* (5 items, Cronbach's alpha = .887), *Involvement* (6 items, two originally from Ethics, Cronbach's alpha = .817), and *Distance* (4 items, Cronbach's alpha = .825).

The use of the Promax rotation was justified by the correlation matrix of the solution (Table 4).

Table 4. Correlation matrix of EFA

Factor Correlation Matrix

Factor	1	2	3
1	1.000	.400	-.437
2	.400	1.000	-.723
3	-.437	-.723	1.000

Extraction Method: Maximum Likelihood.

Rotation Method: Promax with Kaiser Normalization.

Affordances and Use Intentions in the Robot Doctor condition (N=68)

Due to a mishap in the online survey, the experiential variables *Affordances* and *Use Intentions* were measured for a mere sub sample of participants in the Robot Doctor study (*n*=68). To assess the factor structure of these two variables, we executed EFA analysis (Maximum likelihood estimation) with Promax rotation, expecting 2 factors. Bartlett's Test of Sphericity was significant (χ^2 = 1013.44, p < .001) and the Kaiser-Meyer-Olkin Measure of Sampling Adequacy adequately high at .924 (should be > .600). Both these measures indicated that executing EFA made sense for these data. Model fit was good (χ^2= 148.67, p < .001, RMSEA=.100).

The pattern matrix in Table 5 shows the loadings of each item onto each factor. Loadings < .3 were removed for clarity.

Table 5. EFA factor loadings with Promax rotation: *Affordances, Use Intentions*

Pattern Matrix[a]

		Factor	
		1	2
UseA	Ik zou deze robot-dokter echt willen gebruiken.	.575	.358
UseB	Ik zou me laten helpen door de robot-dokter.	.693	
UseC	Recoded: Ik zou de robot-dokter negeren.	.851	
UseD	Recoded: Ik zou de robot-dokter best willen overslaan.	.934	
UseE	Recoded: Ik zou de robot-dokter wegsturen.	.973	
UseF	Ik zou de robot-dokter best nog een keer willen spreken.	.587	.335
AffA	Recoded: Deze robot-dokter is incapabel.	.413	.437
AffB	Deze robot-dokter is kundig.		.758
AffC	Recoded: Deze robot-dokter is dom.		.612
AffD	Deze robot-dokter is handig.		.760

AffE	Recoded: Deze robot-dokter is onhandig.	.450	
AffF	Deze robot-dokter is capabel.		1.040
AffG	Recoded: Deze robot-dokter is klunzig.	.451	
AffH	Recoded: Deze robot-dokter is knullig.	.432	
AffI	Deze robot-dokter is vaardig.		.901
AffJ	Deze robot-dokter is intelligent.		.653

Extraction Method: Maximum Likelihood.
Rotation Method: Promax with Kaiser Normalization.
a. Rotation converged in 3 iterations.

Whereas some of the items clearly loaded onto their respective factors, other showed cross-loadings, whereas yet other loaded on a factor they theoretically did not belong to. UseA and UseF both had a relatively low cross loading on the *Affordances* factor. Because the cross loading on their original factor was higher than the cross loading on the other, we kept both items in the *Use Intentions* factor. In contrast, the item AffA had an almost equal factor loading on both factors. Because of this, we decided to exclude this item from further analyses.

Finally, three *Affordances* items (E, G, and H) had a very low loading on their own factor (< .30), and a higher loading on the *Use Intentions* factor. These three items all had to do with how inadequate the doctor was. While it could be hypothesized that participants group a doctor's clumsiness together with their desire (or lack thereof) to "use" the doctor again in the future, this did not seem like an obvious explanation. Thus, we decided that it was safer to exclude these three items from further analyses.

This left us with the following scales: *Use Intentions* (6 items, Cronbach's alpha = .930), and *Affordances* (6 items, Cronbach's alpha = .948).

The use of the Promax rotation was justified by the correlation matrix of the solution (Table 6).

Table 6. Correlation matrix of EFA

Factor Correlation Matrix

Factor	1	2
1	1.000	.766
2	.766	1.000

Extraction Method: Maximum Likelihood.
Rotation Method: Promax with Kaiser Normalization.

Table 7. Descriptive statistics

Doctor		Human doctor (N = 115)												Robot doctor (N = 134)									
Language		Negation (n = 59)						Affirmation (n = 56)						Negation (n = 68)						Affirmation (n = 66)			
Frame		Negative (n = 32)			Positive (n = 27)			Negative (n = 26)			Positive (n = 30)			Negative (n = 33)			Positive (n = 35)			Negative (n = 31)			Positive (n = 35)	
		n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Familiar Bekhterev?	Yes	3			5			6			4			5			2			5			4		
	No	29			22			20			26			28			33			26			31		
Experience with doctor/patient conv.	Yes	21			18			19			24			21			17			17			20		
	No	11			9			7			6			12			18			14			15		
Sex	Male	9			8			7			9			13			14			13			12		
	Female	23			19			19			21			20			21			18			23		
Age			28.09	13.74		26.74	13.37		27.00	11.72		26.87	11.78		28.52	13.37		29.60	13.63		28.26	11.27		25.86	9.81
Doctor Evaluation			3.65	1.12		3.27	1.19		3.31	1.18		4.03	1.37		4.32	1.44		4.25	1.23		4.06	1.34		3.82	1.37
Message Evaluation			4.00	1.66		2.56	1.51		4.05	1.28		3.60	1.48		4.81	1.51		5.05	1.40		5.25	1.28		4.87	1.31
Expected Quality of Life			4.98	1.28		3.63	1.40		4.42	1.24		3.37	1.14		5.06	1.49		4.61	1.63		3.18	1.41		3.16	1.16
Medical Adherence			5.50	1.34		4.98	1.40		5.31	1.52		5.32	1.53		5.86	1.1		5.73	1.07		5.39	1.21		5.46	1.2
Ethics															5.13	0.87		4.94	0.85		4.94	1.03		4.59	1.08
Involvement															2.97	0.86		2.90	0.84		2.62	0.93		2.72	0.92
Distance															3.97	1.37		4.19	1.14		4.68	1.01		4.44	1.35
Language														Negation (n = 35)						Affirmation (n = 33)					
Frame														Negative (n = 16)			Positive (n = 19)			Negative (n = 16)			Positive (n = 17)		
														M	*SD*		*M*	*SD*		*M*	*SD*		*M*	*SD*	
Affordances														3.68	0.79		3.48	0.70		3.36	0.81		3.30	0.86	
Use Intentions														3.26	0.35		2.99	0.42		3.16	0.47		3.39	0.46	

Note. These are background variables that are available across samples.

Main Analyses

Message-related outcome variables

Preliminary Analyses. Before testing our hypothesis, we tested whether the experimental condition assignment did not result in any differences in background information between condition cells. An ANOVA with age as the dependent variable showed no main or interaction effects of the condition variables.

We created a variable that contained the combination of our three condition variables (Doctor, Frame, Language) and used this variable to explore differences in categorical variables with a Chi-square test. There were no significant dependencies between conditions and familiarity with Bekhterev's disease (χ^2= 5.24, p = .630, φ = .145), experience with doctor/patient conversations (χ^2= 9.64, p = .210, φ = .197), or gender (χ^2= 3.37, p = .849, φ = .116).

Table 7 shows the descriptive statistics of the sample, including the Burgers data. Because there are no differences between the conditions on the background variables, we will not include them in further analyses.

MANOVA. To test our hypotheses, we performed a 2 (Doctor: Human vs. Robot) x 2 (Language: Negation vs. Affirmation) x 2 (Frame: Negative vs. Positive) between-subjects MANOVA. *Message Evaluation, Doctor Evaluation, Medication Adherence,* and *Expected Quality of Life* were the dependent variables.

Multivariate Results. The multivariate results showed that all main effects and two two-way interaction effects (Doctor * Language and Doctor * Framing) were significant, see Table 8.

Table 8. Multivariate results 2x2x2 MANOVA

	Wilk's λ	F	*df hypo*	*df error*	p	Partial η^2
Doctor	0.791	15.74	4	238	< .001	.209
Language	0.853	10.24	4	238	< .001	.149
Frame	0.890	7.33	4	238	< .001	.110
Doctor * Language	0.940	3.78	4	238	.005	.060
Doctor * Frame	0.918	5.33	4	238	< .001	.082
Language * Frame	0.987	0.76	4	238	.555	.013
Doctor * Language * Frame	0.973	1.66	4	238	.161	.027

Univariate Results. Table 9 shows the univariate results, in which *Doctor Evaluation, Message Evaluation*, and *Medical Adherence* showed a univariate effect of Doctor, whereas *Expected Quality of Life* was not affected by Doctor. For all three dependents, the Human Doctor scored lower than the Robot Doctor: *Doctor Evaluation* was higher for the Robot Doctor (M = 4.11, SE = .11, 95% CI 3.89 – 4.33) than for the Human Doctor (M = 3.57, SE = .12, 95% CI 3.32 – 4.80). *Message Evaluation* was higher for the Robot Doctor (M = 4.99, SE = .12, 95% CI 4.75 – 5.24) than the Human Doctor (M = 3.55, SE = .14, 95% CI 3.29 – 3.82). *Medical Adherence* was higher for the Robot Doctor (M = 5.61, SE = .11, 95% CI 5.39 – 5.83) than the Human Doctor (M = 5.28, SE = .12, 95% CI 5.04 – 5.51).

Language only had a univariate effect on *Expected Quality of Life,* where Affirmation resulted in lower scores (M = 3.53, SE = .12, 95% CI 3.29 – 3.78) than Negation (M = 4.57, SE = .12, 95% CI 4.33 – 4.81).

Framing had an effect on both *Expected Quality of Life* and *Message Evaluation.* In both cases, positive framing resulted into a lower score. *Expected Quality of Life* was lower in a Positive Frame (M = 3.69, SE = .12, 95% CI 3.45 – 3.93) than in a Negative Frame (M = 4.41, SE = .12, 95% CI 4.17 – 4.66).[1] *Message Evaluation* was lower in a Positive Frame (M = 4.02, SE = .13, 95% CI 3.77 – 4.27) than in a Negative Frame (M = 4.53, SE = .13, 95% CI 4.27 – 4.78).

The interaction effect of Doctor and Language had a significant effect on *Expected Quality of Life.* Pairwise comparisons showed that there was no effect of Language for the Human Doctor (ΔM = 0.41, SE = .25, p = .106, 95% CI = -0.09 – 0.92), whereas there was an effect of Language for the Robot Doctor (ΔM = 1.67, SE = .24, p < .001, 95% CI = 1.21 – 2.13). Figure 1 exhibits a visual representation of this interaction effect: If a Robot Doctor uses Negation in its message, the *Expected Quality of Life* is on average higher than when the Robot Doctor uses Affirmation in its message. In contrast, there is no effect of Language on *Expected Quality of Life* for the Human Doctor, the line is almost horizontal.

The interaction effect of Doctor and Frame had a significant effect on both *Expected Quality of Life* and *Message Evaluation.* In contrast to the interaction effect with Language, pairwise comparisons for *Expected Quality of Life* showed that there was a significant effect of Framing for the Human Doctor (ΔM = 1.21, SE = .25, p <

[1] This is the reverse result of the Burgers et al. study, which probably had a scale-analysis issue (see section *Reanalysis of Burgers et al. (2012)*).

.001, 95% *CI* = 0.71 – 1.71), but not for the Robot Doctor (*ΔM* = 0.23, *SE* = .24, *p* = .322, 95% *CI* = -0.23 – 0.70). Figure 2 displays a visual representation of this interaction effect. The figure shows that if a Human Doctor uses Negative Frames, the *Expected Quality of Life* is on average higher than when the Human Doctor uses Positive Frames. By contrast, there is no effect of Frame on *Quality of Life* for the Robot Doctor, the line is almost horizontal.

With regard to the interaction effect of Doctor and Frame on *Message Evaluation,* a similar pattern emerged. Pairwise comparisons showed that there was a significant effect of Frame for the Human Doctor (*ΔM* = 0.95, *SE* = .27, *p* = .001, 95% *CI* = 0.42 – 1.48), but not for the Robot Doctor (*ΔM* = 0.07, *SE* = .25, *p* = .777, 95% *CI* = -0.42 – 0.56). Figure 3 provides a visual representation of this interaction effect: If a Human Doctor used Negative Framing, the *Message Evaluation* on average was higher than when the Human Doctor used Positive Framing. In contrast, there was no effect of Frame on *Message Evaluation* for the Robot Doctor, the line is almost horizontal. However, as we know from the main effect of Doctor, *Message Evaluation* across Framing conditions was higher for the Robot Doctor than for the Human Doctor.

Table 9. Univariate results 2x2x2 MANOVA: significant multivariate results

IV	DV	*df*	*F*	*p*	Partial η^2
Corrected Model	Doctor Evaluation	7	2.839	.007	.076
	Message Evaluation	7	11.629	.000	.252
	Expected Quality of Life	7	11.379	.000	.248
	Medical Adherence	7	1.329	.237	.037
Intercept	Doctor Evaluation	1	2188.197	.000	.901
	Message Evaluation	1	2175.424	.000	.900
	Expected Quality of Life	1	2190.539	.000	.901
	Medical Adherence	1	4376.624	.000	.948
Doctor	Doctor Evaluation	1	11.182	.001	.044
	Message Evaluation	1	61.853	.000	.204
	Expected Quality of Life	1	.324	.570	.001
	Medical Adherence	1	4.087	.044	.017

Language	Doctor Evaluation	1	.168	.682	.001
	Message Evaluation	1	3.415	.066	.014
	Expected Quality of Life	1	36.162	.000	.130
	Medical Adherence	1	.845	.359	.003
Framing	Doctor Evaluation	1	.003	.954	.000
	Message Evaluation	1	7.727	.006	.031
	Expected Quality of Life	1	17.267	.000	.067
	Medical Adherence	1	.762	.384	.003
Doctor * Language	Doctor Evaluation	1	2.754	.098	.011
	Message Evaluation	1	1.306	.254	.005
	Expected Quality of Life	1	13.199	.000	.052
	Medical Adherence	1	1.832	.177	.008
Doctor * Framing	Doctor Evaluation	1	.975	.324	.004
	Message Evaluation	1	5.734	.017	.023
	Expected Quality of Life	1	7.884	0.005	0.032
	Medical Adherence	1	.456	.500	.002
Error	Doctor Evaluation	241			
	Message Evaluation	241			
	Expected Quality of Life	241			
	Medical Adherence	241			
Total	Doctor Evaluation	249			
	Message Evaluation	249			
	Expected Quality of Life	249			
	Medical Adherence	249			
Corrected Total	Doctor Evaluation	248			
	Message Evaluation	248			
	Expected Quality of Life	248			
	Medical Adherence	248			

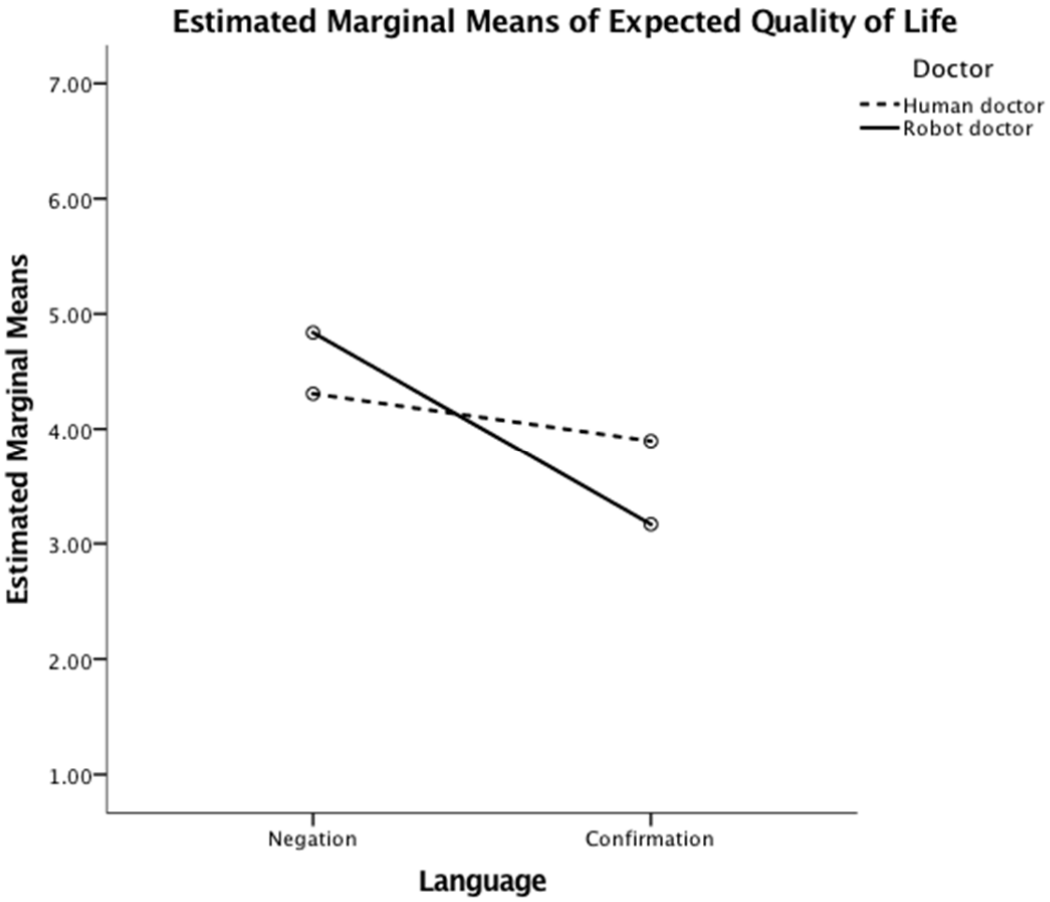

Figure 1. Interaction Effect of Doctor * Language on *Expected Quality of Life*

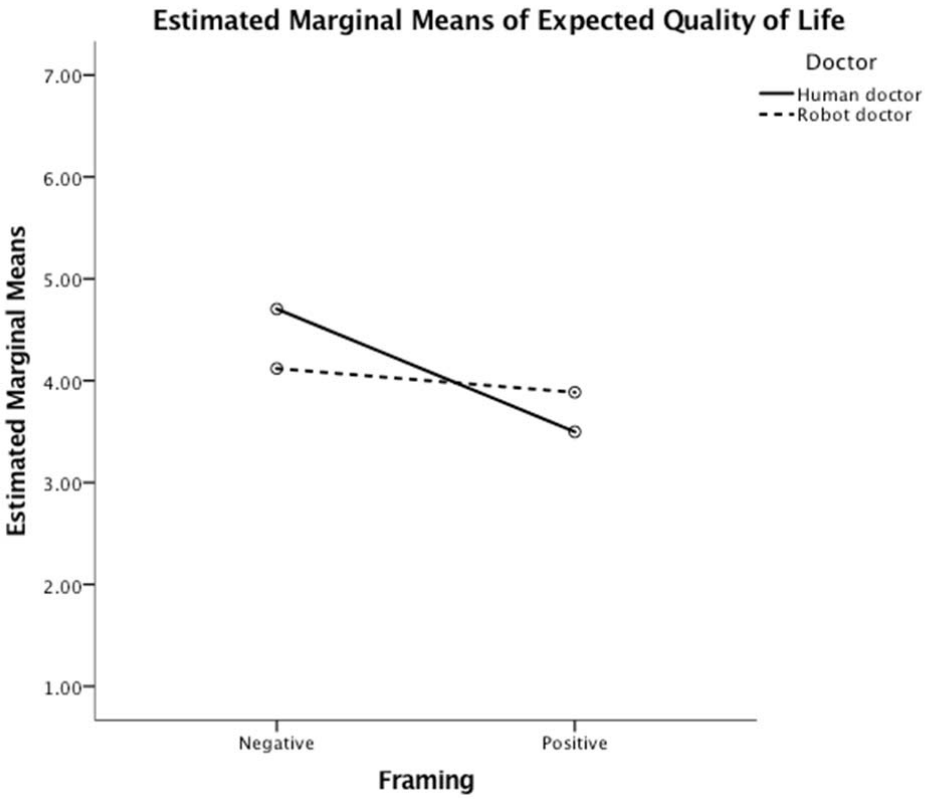

Figure 2. Interaction Effect of Doctor * Frame on *Expected Quality of Life*

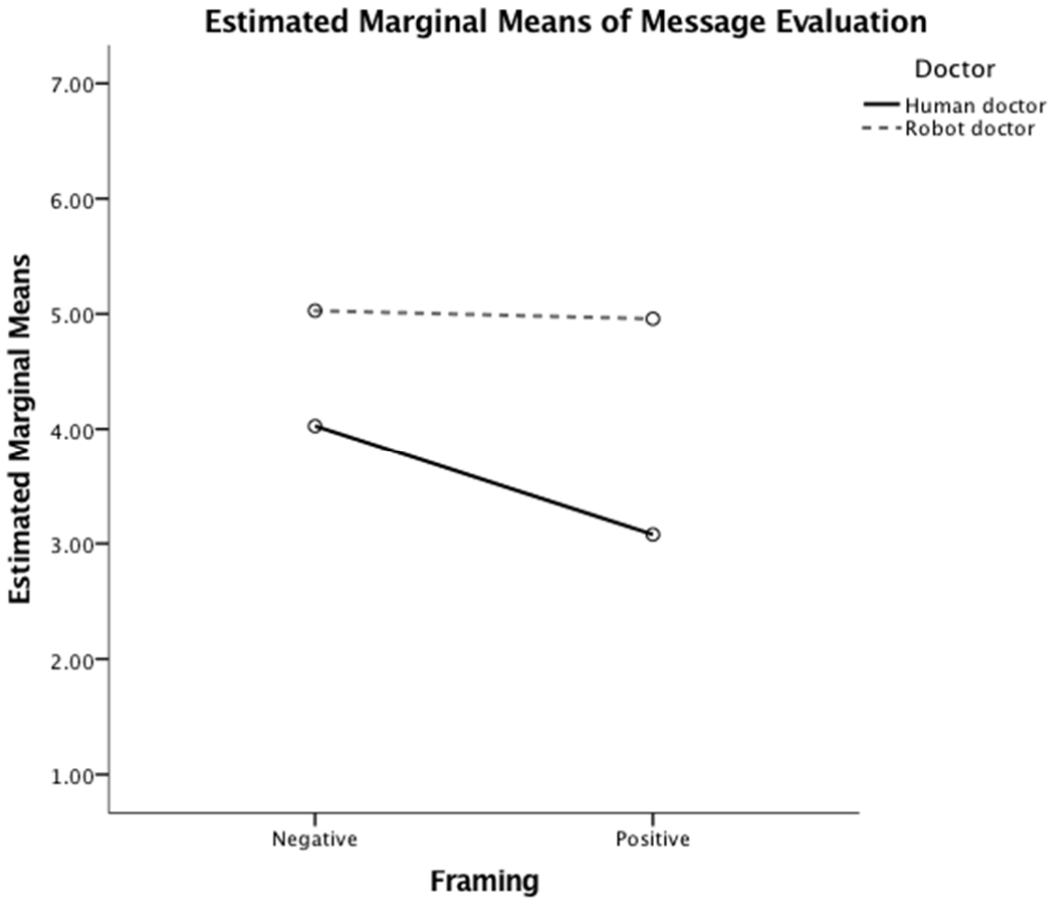

Figure 3. Interaction effect of Doctor * Frame on *Message Evaluation*

Reanalysis of Burgers et al. (2012)

Particularly the data for *Expected Quality of Life* showed different results in the Burgers et al. study as compared to the current results. Burgers et al report: "Participants in the condition with positive framing were more positive about the expected quality of life (*M* = 4.29, *SD* = 1.59) than participants in the condition with negative framing (expected quality of life: *M* = 2.96, *SD* = 1:32)." For all other variables, they did not find differences or they found the reverse that negative frames led to more positive evaluations (mainly because 'not good' was valued higher than 'not bad'), just like we did. In the current study, we also found that negative frames yielded higher scores than positive frames, except that it also applied to *Expected Quality of Life*. This is the opposite of Burgers et al., where positive frames scored higher on *Expected Quality of Life* than negative ones. One might think that this reversal is due to interference by the robot data, but if we look at the isolated human-

doctor data in our study (which are a reanalysis of Burgers et al.), then we still find the reverse of the original Burgers et al. analysis that *Expected Quality of Life* is higher in negative frames than in positive ones. What can the matter be?

First off, we checked whether we recoded the contra-indicative items properly, which was the case. Then we contacted the authors of the Burgers et al. study to find out whether they made a mistake, which was also not the case. Then we noticed that the Burgers et al. study checked on convergent validity of the measurement scales but not on the divergent validity, which we did by running factor analyses. Hence, we used slightly different scales from the Burgers et al. study, which probably is responsible for the different results.

Full Replication Approach. In addition, then, we did a full replication of Burgers et al. (2012) and kept the items of the message-related outcome variables identical. This left us with the following scales: *Message Evaluation* (4 items, Cronbach's alpha = .683), *Doctor Evaluation* (5 items, Cronbach's alpha = .884), *Expected Quality of Life* (2 items, Cronbach's alpha = .111), *Medical Adherence* (2 items, Cronbach's alpha = .842).

Because in the replication, the measurement scales were composed differently from our study, the mean scale values were different as well. In Table 10, the descriptives of the replication used in the reanalysis are tabulated.

The pattern of results of the multivariate analysis in the replication was the same as in our approach (see Table 8), although the absolute values differed, of course (Table 11). Univariate effects also showed similar patterns as before albeit with different values: *Message Evaluation* was higher for the Robot Doctor (M = 4.76, SE = .10, 95% CI 4.56 – 4.95) than the Human Doctor (M = 3.73, SE = .11, 95% CI 3.52 – 3.94) and Language had a univariate effect on *Expected Quality of Life*, where Affirmation resulted in lower scores (M = 3.30, SE = .10, 95% CI 3.10 – 3.50) than Negation (M = 4.15, SE = .10, 95% CI 3.95 – 4.35). Framing again affected *Message Evaluation*, showing that a Positive Frame yielded lower scores (M = 3.99, SE = .10, 95% CI 3.79 – 4.19) than Negative Frames (M = 4.50, SE = .10, 95% CI 4.29 – 4.70).

But here was the surprise: The problematic results of Frame on *Expected Quality of Life* turned out to be *not* significant (!) in this full replication approach to the Burgers et al. (2012) data of their Experiment 2.

Table 10. Descriptive statistics replicating the original approach of Burgers et al. (2012). The differences with our study are marked grey.

Doctor	Human doctor (N = 115)												Robot doctor (N = 134)											
Language	Negation (n = 59)						Affirmation (n = 56)						Negation (n = 68)						Affirmation (n = 66)					
Frame	Negative (n = 32)			Positive (n = 27)			Negative (n = 26)			Positive (n = 30)			Negative (n = 33)			Positive (n = 35)			Negative (n = 31)			Positive (n = 35)		
	n	M	SD	n	M	SD	n	M	SD	n	M	SD	n	M	SD	n	M	SD	n	M	SD	n	M	SD
Doctor Evaluation		3.65	1.12		3.27	1.19		3.31	1.18		4.03	1.37		4.32	1.44		4.25	1.23		4.06	1.34		3.82	1.37
Message Evaluation		4.17	1.31		2.93	1.16		4.21	1.09		3.62	1.14		4.88	1.32		5.00	1.22		4.73	.99		4.41	.83
Expected Quality of Life		4.03	.79		3.48	.70		3.44	1.21		3.66	.56		4.80	1.50		4.28	1.45		3.05	1.17		3.04	1.12
Medical Adherence		5.50	1.34		4.98	1.40		5.31	1.52		5.32	1.53		5.86	1.1		5.73	1.07		5.39	1.21		5.46	1.2

Note. These are background variables that are available across samples.

Table 11. Multivariate results of the 2x2x2 MANOVA in the replication analysis

	Wilk's λ	F	$df\ hypo$	$df\ error$	p	Partial η^2
Doctor	0.824	12.73	4	238	< .001	.176
Language	0.860	9.72	4	238	< .001	.140
Frame	0.937	4.01	4	238	.004	.063
Doctor * Language	0.916	5.43	4	238	< .001	.084
Doctor * Frame	0.942	3.67	4	238	.006	.058
Language * Frame	0.974	1.56	4	238	.185	.026
Doctor * Language * Frame	0.977	1.38	4	238	.243	.023

Again (Table 12), the interaction effect of Doctor and Language had a significant effect on *Expected Quality of Life* but this time also on *Message Evaluation*, although this turned out to be inconsequential after all. With regard to *Expected Quality of Life*, pairwise comparisons showed that there was no significant effect of Language for the Human Doctor (ΔM = 0.20, SE = .21, p = .339, 95% CI = -0.21 – 0.62), but again, there was for the Robot Doctor (ΔM = -1.50, SE = .20, p < .001, 95% CI = -1.88 – -1.12) (Figure 4). With respect to *Message Evaluation*, none of the pairwise comparisons were significant (Table 12), indicating that any mean differences between conditions were small (see Figure 5 for a visual representation).

Table 12. Univariate Results 2x2x2 MANOVA in the replication approach

IV	DV	*df*	*F*	*p*	Partial η^2
Corrected Model	Doctor Evaluation	7	2.839	.007	.076
	Message Evaluation	7	10.920	.000	.241
	Expected Quality of Life	7	9.736	.000	.220
	Medical Adherence	7	1.329	.237	.037
Intercept	Doctor Evaluation	1	2188.197	.000	.901
	Message Evaluation	1	3386.076	.000	.934
	Expected Quality of Life	1	2692.404	.000	.918
	Medical Adherence	1	4376.624	.000	.948
Doctor	Doctor Evaluation	1	11.182	.001	.044
	Message Evaluation	1	49.398	.000	.170
	Expected Quality of Life	1	0.945	.332	.004
	Medical Adherence	1	4.087	.044	.017
Language	Doctor Evaluation	1	0.168	.682	.001
	Message Evaluation	1	0.000	1.000	.000
	Expected Quality of Life	1	35.070	.000	.127
	Medical Adherence	1	0.845	.359	.003
Frame	Doctor Evaluation	1	0.003	.954	.000
	Message Evaluation	1	12.214	.001	.048
	Expected Quality of Life	1	2.181	.141	.009
	Medical Adherence	1	0.762	.384	.003

Doctor * Language	Doctor Evaluation	1	2.754	.098	.011
	Message Evaluation	1	6.270	.013	.025
	Expected Quality of Life	1	20.394	.000	.078
	Medical Adherence	1	1.832	.177	.008
Doctor * Frame	Doctor Evaluation	1	0.975	.324	.004
	Message Evaluation	1	7.924	.005	.032
	Expected Quality of Life	1	0.118	.731	.000
	Medical Adherence	1	0.456	.500	.002
Error	Doctor Evaluation	241			
	Message Evaluation	241			
	Expected Quality of Life	241			
	Medical Adherence	241			
Total	Doctor Evaluation	249			
	Message Evaluation	249			
	Expected Quality of Life	249			
	Medical Adherence	249			
Corrected Total	Doctor Evaluation	248			
	Message Evaluation	248			
	Expected Quality of Life	248			
	Medical Adherence	248			

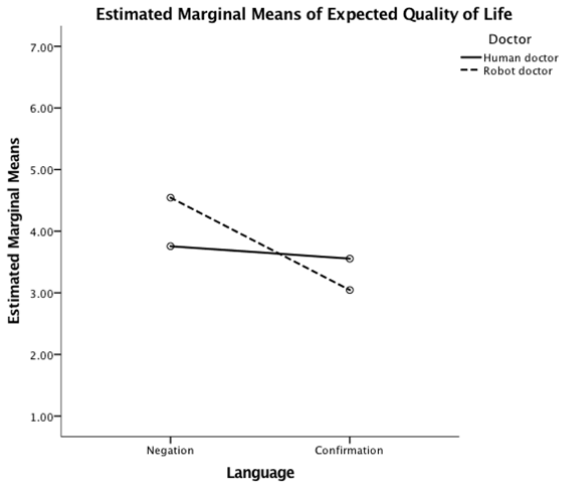

Figure 4. Interaction Doctor * Language on *Expected Quality of Life* in the replication

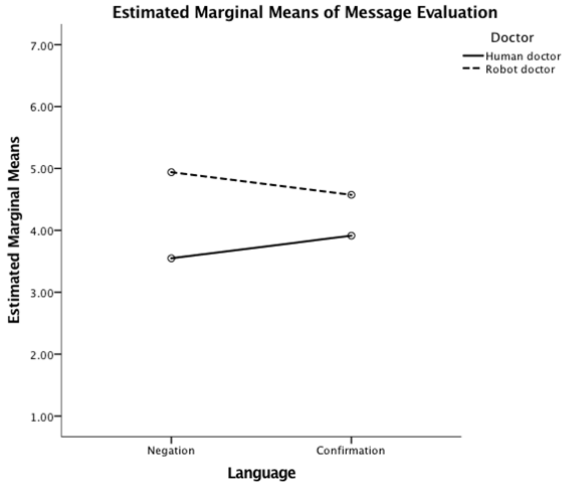

Figure 5. Interaction of Doctor * Language on *Message Evaluation* in the replication

Expected Quality of Life: Single-item Approach. As an extra control, we also analyzed the one item of *Expected Quality of Life* that was central in its scale and never contaminated another scale in the factor analyses: "Op basis van dit gesprek schat ik mijn levenskwaliteit hoog in" ("On the basis of this conversation, I expect my quality of life will be high"). Multivariate analyses of the mean scores to this item (Table 13) showed, as before (Table 8), that the main effects and the two two-way interaction effects were significant (Doctor * Language and Doctor * Framing, Table 14).

Univariate analyses showed again, with slightly different values, that Language had a univariate effect on *Expected Quality of Life;* Affirmation being lower (M = 3.52, SE = .14, 95% CI 2.92 – 3.47) than Negation (M = 3.82, SE = .14, 95% CI 3.56 – 4.09) (Table 15, Figure 6). However, for the first time, we found the result reported by Burgers et al. (2012) that a Positive Frame raised *higher* scores for *Expected Quality of Life* (M = 3.76, SE = .14, 95% CI 3.49 – 4.03) than a Negative Frame (M = 3.26, SE = .12, 95% CI 4.17 – 4.66) (Table 15). Point is, we could reproduce Burgers et al. by using a single item and not by pure replication, using the exact same scale they used.

Table 13. Descriptive statistics in the single-item approach. The differences with our study are marked grey.

Doctor	Human doctor (N = 115)												Robot doctor (N = 134)											
Language	Negation (n = 59)						Affirmation (n = 56)						Negation (n = 68)						Affirmation (n = 66)					
Frame	Negative (n = 32)			Positive (n = 27)			Negative (n = 26)			Positive (n = 30)			Negative (n = 33)			Positive (n = 35)			Negative (n = 31)			Positive (n = 35)		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*N*	*M*	*SD*	*n*	*M*	*SD*
Doctor Evaluation		3.65	1.12		3.27	1.19		3.31	1.18		4.03	1.37		4.32	1.44		4.25	1.23		4.06	1.34		3.82	1.37
Message Evaluation		4.00	1.66		2.56	1.51		4.05	1.28		3.60	1.48		4.81	1.51		5.05	1.40		5.25	1.28		4.87	1.31
Expected Quality of Life		2.78	1.31		3.74	1.87		2.73	1.54		4.27	1.60		4.58	1.77		4.20	1.61		2.94	1.32		2.83	1.18
Medical Adherence		5.50	1.34		4.98	1.40		5.31	1.52		5.32	1.53		5.86	1.1		5.73	1.07		5.39	1.21		5.46	1.2

Note. These are background variables that are available across samples.

Table 14. Multivariate results 2x2x2 MANOVA for single item *Expected Quality of Life*

	Wilk's λ	*F*	*df hypo*	*df error*	*p*	Partial η^2
Doctor	0.791	15.75	4	238	< .001	.209
Language	0.923	4.96	4	238	.001	.077
Frame	0.923	4.96	4	238	.001	.077
Doctor * Language	0.922	5.03	4	238	.001	.078
Doctor * Frame	0.891	7.26	4	238	< .001	.109
Language * Frame	0.988	0.72	4	238	.582	.012
Doctor * Language * Frame	0.973	1.65	4	238	.163	.027

Also in the single-item approach, the interaction effect of Doctor and Frame had a significant effect on *Expected Quality of Life.* Pairwise comparisons indicated a significant effect of Frame for the Human Doctor (ΔM = 1.25, SE = .29, p < .001, 95% CI = 0.68 – 1.81) and not for the Robot Doctor (ΔM = -0.24, SE = .27, p = .363, 95% CI = -0.76 – 0.28); also see Figure 7.

Table 15. Univariate results 2x2x2 MANOVA: single-item *Expected Quality of Life*

IV	DV	*df*	*F*	*p*	Partial η^2
Corrected Model	Doctor Evaluation	7	2.839	.007	.076
	Message Evaluation	7	11.629	.000	.252
	Expected Quality of Life	7	8.055	.000	.190
	Medical Adherence	7	1.329	.237	.037
Intercept	Doctor Evaluation	1	2188.197	.000	.901
	Message Evaluation	1	2175.424	.000	.900
	Expected Quality of Life	1	1294.694	.000	.843
	Medical Adherence	1	4376.624	.000	.948
Doctor	Doctor Evaluation	1	11.182	.001	.044
	Message Evaluation	1	61.853	.000	.204
	Expected Quality of Life	1	1.712	.192	.007
	Medical Adherence	1	4.087	.044	.017
Language	Doctor Evaluation	1	.168	.682	.001
	Message Evaluation	1	3.415	.066	.014
	Expected Quality of Life	1	10.578	.001	.042
	Medical Adherence	1	.845	.359	.003
Framing	Doctor Evaluation	1	.003	.954	.000
	Message Evaluation	1	7.727	.006	.031
	Expected Quality of Life	1	6.662	.010	.027
	Medical Adherence	1	.762	.384	.003
Doctor * Language	Doctor Evaluation	1	2.754	.098	.011
	Message Evaluation	1	1.306	.254	.005
	Expected Quality of Life	1	19.997	.000	.077
	Medical Adherence	1	1.832	.177	.008

Doctor * Framing	Doctor Evaluation	1	.975	.324	.004
	Message Evaluation	1	5.734	.017	.023
	Expected Quality of Life	1	14.584	.000	.057
	Medical Adherence	1	.456	.500	.002
Error	Doctor Evaluation	241			
	Message Evaluation	241			
	Expected Quality of Life	241			
	Medical Adherence	241			
Total	Doctor Evaluation	249			
	Message Evaluation	249			
	Expected Quality of Life	249			
	Medical Adherence	249			
Corrected Total	Doctor Evaluation	248			
	Message Evaluation	248			
	Expected Quality of Life	248			
	Medical Adherence	248			

Figure 6. Interaction Doctor * Language on single-item *Expected Quality of Life*

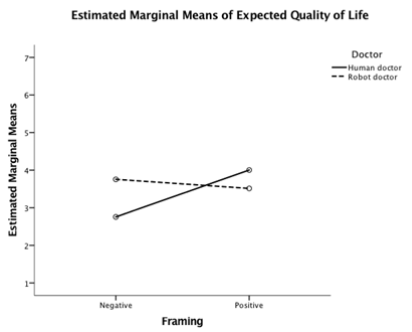

Figure 7. Interaction Doctor * Framing on single-item *Expected Quality of Life*

Table 16 provides a comparison of the three approaches (divergent validity through factor analysis, a full replication of Burgers et al., and the single-item approach). There are but a few differences among the three but one of them is crucial. In the full replication, the effects of Frame on *Expected Quality of Life* are not significant and the Doctor*Language interaction effect on *Message Evaluation* only in the general ANOVA, not in the pairwise contrasts.

The crucial point that started this reanalysis of the Burgers et al. data is that the pattern they found (Positive Frame scores higher on *Expected Quality of Life* than Negative Frame) can only be replicated for the human doctor using a single item, not with the exact scale they used, which rendered insignificant results (Table 16, in red). This led us to think that we should stick to our own approach and conclude that Negative Frames are preferred for each of the four outcome variables.

Table 16. Overview of results for three approaches to the analysis. R = Robot doctor, H = Human doctor

IV	**DV**	**Version 1**	**Version 2**	**Version 3**
Doctor	Doctor Evaluation	R > H	R > H	R > H
	Message Evaluation	R > H	R > H	R > H
	Expected Quality of Life	*ns*	*ns*	*ns*
	Medical Adherence	R > H	R > H	R > H
Language	Doctor Evaluation	*ns*	*ns*	*ns*
	Message Evaluation	*ns*	*ns*	*ns*
	Expected Quality of Life	N > C	N > C	N > C
	Medical Adherence	*ns*	*ns*	*ns*

Frame	Doctor Evaluation	*ns*	*ns*	*ns*
	Message Evaluation	N > P	N > P	N > P
	Expected Quality of Life	N > P	N > P	*ns*
	Medical Adherence	*ns*	*ns*	*ns*
Doctor * Language	Doctor Evaluation	*ns*	*ns*	*ns*
	Message Evaluation	*ns*	*ns*	Anova sig, pairwise *ns*.
	Expected Quality of Life	H: *ns*; R: N > C	H: *ns*; R: N > C	H: *ns*; R: N > C
	Medical Adherence	*ns*	*ns*	*ns*
Doctor * Frame	Doctor Evaluation	*ns*	*ns*	*ns*
	Message Evaluation	H: N > P; R: *ns*	H: N > P; R: *ns*	H: N > P; R: *ns*
	Expected Quality of Life	H: N > P; R: *ns*	H: P > N; R: *ns*	*ns*
	Medical Adherence	*ns*	*ns*	*ns*

Version 1: Scales based on Factor Analysis, two-item *Expected QoL* scale
Version 2: Scales based on Factor Analysis, one-item *Expected QoL* scale
Version 3: Exact replication of scales used in Burgers et al. (2012)

Bayesian Statistics

For a non-technical, introductory article on Bayesian statistics, please see Van de Schoot and Depaoli (2014). In conventional (frequentist) statistical techniques, data are always compared to a null hypothesis. This works fine when one has no idea of what could be going on in the data. However, this is not often the case. Researchers have theories, expectations, and previous studies to build their case on (prior knowledge). Bayesian statistics allows prior knowledge (the prior) to support the estimation of a model and to test hypotheses. Using this method, one can build on earlier research, instead of starting from scratch every time anew.

Moreover, frequentist statistics assume that in the population there exists but one true population parameter (fixed). In Bayesian statistics, all unknown parameters can be defined by a probability distribution. Thus, Bayesian statistics do not result in

a point estimate, but rather in an interval with a certain probability that the true coefficient is part of that interval.

Bayesian analysis exists of three parts: the prior distribution, the data (the likelihood), and the posterior distribution. The posterior distribution is a combination of the prior distribution and the data, an updated understanding of the theory under question.

For a more technical introduction to Bayesian statistics, please see Gelman, Carlin, Stern, and Rubin (2014) and Van de Schoot, Kaplan, Denissen, Asendorpf, Neyer, & Aken (2014).

Bayes factors

Human Doctors = Robot Doctors? Based on our theoretical framework, we were interested in figuring out the amount of evidence that exists for a Human and Robot doctor to be perceived "the same" when they follow the same rules for communication frames and language. We hypothesized that participants' evaluations of our four outcome measures would be higher with affirmative language and positive frames. For the interaction effect of Frame and Language, we expected that positive frame and affirmative language would lead to the highest scores, followed by negative frame with affirmative language, and positive frame with negation language. The lowest scores were expected for the negative frame with negation language.

Bayesian hypothesis testing in JASP. A relatively new program that allows the comparison of various models (not effects) to each other while using Bayes Factors is JASP (Love, Selker, Verhagen, et al., 2015; Morey & Rouder, 2015; Rouder, Morey, Speckman, & Province, 2012). In an attempt to become a free alternative to SPSS, the program is still under development so that today, it is not possible to perform a MANOVA within JASP. Thus, we will perform 4 separate ANOVAs, each testing one of our outcome variables.

It is important to understand that JASP does not return a Bayes Factor for each effect in a full model (a model with all main and interaction effects). Instead, it returns a Bayes Factor for each model, building from a Null model (no predictors) to a model with all main effects and interactions included. For each step, JASP compares the current model to the original Null model and computes a Bayes Factor based on the difference in model fit of the two models. JASP can produce two types of Bayes Factors, one that quantifies evidences in favor of the Null model as compared to the

Alternative model (BF_{01}), and another that quantifies the opposite evidence (in favor of the Alternative model as compared to the Null Model: BF_{10}).

Doctor Evaluation. Table 17 shows partial output of the Bayesian ANOVA performed in JASP. Instead of including all steps in the model-building sequence, we limited the table to "traditional steps". First, all main effects are tabulated on their own, after which all main effects are included in one model. Next, we added all 2-way interaction effects, and finally, we incorporated the 3-way interaction effect. Table 17 shows both types of Bayes Factors.

We hypothesized that there would be no main effect of Doctor, i.e. that the BF in favor of the null model compared to a model including the Doctor main effect would be high. However, if we look at the BF_{01} for the model including only the main effect of Doctor, we see that it is < 1, indicating that there is no evidence in favor of the Null model. This is further supported by a high BF_{10}, which shows that there is actually evidence that a model with the Doctor main effect is better at explaining Doctor Evaluation scores than a model with no predictors.

What is more, the model including only the Doctor main effect has the highest BF_{10} of all models tested (this also includes models tested but not reported in Table 17). Compared to a model with all main effects, the model with just the Doctor main effect was preferred by a Bayes Factor of 49.09 (BF_Doctor / BF_AllMain). This preference became even more pronounced when the Doctor-only model was compared to a model including all 2-way interactions as well as the 3-way interaction (Bayes Factor = 416.68).

Thus, it seems that *Doctor Evaluation* scores are best explained by who the doctor is (human or robot), whereas Language and Frame play no role. These results are in agreement with the frequentist results. *Doctor Evaluation* was higher for the Robot Doctor (*M* = 4.11, *SE* = .11, 95% *CI* 3.89 – 4.33) than the Human Doctor (*M* = 3.57, *SE* = .12, 95% *CI* 3.32 – 4.80). Thus, the hypothesis that human and robot doctors received equal *Doctor Evaluations* was not confirmed. Furthermore, the hypothesis that Language and Frame play a decisive role in *Doctor Evaluation* could also not be confirmed.

Table 17. Bayesian ANOVA results for Doctor Evaluation

No.	Included effects	BF_{01}	BF_{10}

1.	Doctor	0.059	17.084
2.	Language	6.395	0.156
3.	Frame	7.190	0.139
4.	All main effects	2.874	0.348
5.	No. 4. + all 2-way interactions	35.852	0.028
6.	No. 5. + 3-way interaction	24.130	0.041

Message Evaluation. Table 18 shows partial output of the Bayesian ANOVA performed in JASP. Instead of including all steps in the model-building sequence, we limited the table to "traditional steps" plus the step with the highest BF_{10}. First, all main effects are shown on their own, then all main effects compacted into one model. Subsequently, the model with the highest BF_{10} was included (Doctor and Frame main effects and their interaction). Next, we added all 2-way interaction effects, and finally, we included the 3-way interaction effect. Table 18 shows both types of Bayes Factors.

Again we hypothesized that there would be no main effect of Doctor, i.e. that the BF in favor of the null model compared to a model including the Doctor main effect would be high. However, if we look at the BF_{01} for the model including only the main effect of Doctor, we see that it was < 1, indicating that there was no evidence in favor of the Null model. This is further supported by a high BF_{10}, showing that there is actually evidence that a model with the Doctor main effect was better at explaining *Message Evaluation* scores than a model with no predictors.

While the BF_{10} for the All-main-effects-model was very high, there was one model that resulted in an even higher BF_{10}: A model that incuded the main effect of Doctor and Frame and their interaction. Compared to the All-main-effects-model, this model was preferred by a Bayes Factor of 4.42.

Thus, it seems that *Message Evaluation* scores were best explained by who the doctor was (human or robot), how the message was framed (positive or negative), and by the interaction between these two factors. Language did not seem to play a role. These results are in agreement with the frequentist results. Again, the hypothesis that human and robot perform about equally well could not be confirmed. Furthermore, the interaction effect of Frame and Doctor revealed that, for human doctors, it was negative framing, not positive framing, that led to higher *Message Evaluation* scores. For robot doctors, *Message Evaluation* was not affected by Frame (see Figure 3). This is the opposite of what was expected based on the theory on humans. Finally,

hypotheses regarding the influence of Language on *Message Evaluation* could not be confirmed either.

Table 18. Bayesian ANOVA results for *Message Evaluation*

No.	Included effects	BF_{01}	BF_{10}
1.	Doctor	1.35E-10	7.40E+09
2.	Language	2.922	0.342
3.	Frame	1.021	0.979
4.	All main effects	1.20E-10	8.36E+09
5a.	Doctor + Frame + Doctor*Frame	2.71E-11	3.69E+10
5b.	No. 4. + all 2-way interactions	6.19E-10	1.62E+09
6.	No. 5. + 3-way interaction	3.48E-10	2.88E+09

Medical Adherence. Table 19 lists partial output of the Bayesian ANOVA performed in JASP and "traditional steps" plus the step with the highest BF_{10} in the model-building sequence. First, all main effects are listed on their own, then all main effects comprised into one model. Subsequently, the model with the highest BF_{10} was included (Doctor and Frame main effects and their interaction). Next, we added all 2-way interaction effects, and finally, we included the 3-way interaction effect. Table 19 has both types of Bayes Factors.

As before, we hypothesized that there would be no main effect of Doctor, i.e. that the BF in favor of the null model compared to a model including the Doctor main effect would be high. If we look at the BF_{01} for the model including only the main effect of Doctor, we see that it is 1.128, indicating that there was some evidence in favor of the Null model. This was further supported by a low BF_{10}. Thus, it seems that who the doctor was (human or robot) did not matter, on its own, in predicting *Medical Adherence* scores.

None of the other BF_{10} reached a level higher than 1. Instead, adding more and more predictors to the model resulted in a Bayes Factor that kept declining.

Thus, it seems that *Medical Adherence* scores were not explained by Doctor, Language, or Frame. These results are not in agreement with the frequentist results. There we found that *Medical Adherence* was significantly higher for the Robot

Doctor (*M* = 5.61, *SE* = .11, 95% *CI* 5.39 – 5.83) than for the Human Doctor (*M* = 5.28, *SE* = .12, 95% *CI* 5.04 – 5.51). This discrepancy can be explained by the relatively small mean difference of 0.33 points between the groups combined with a relatively large sample size. Using frequentist methods with a large sample size can result in spurious significant effects that are not necessarily meaningful. Bayesian estimation can lead to decreases in BF with increasing sample size, if the mean difference is not big and specific (low variance within groups) enough to be meaningful (Konijn, Van de Schoot, Winter, & Ferguson, 2015).

Thus, basing our conclusion on the Bayesian analysis, we can carefully confirm the hypothesis that for *Medical Adherence* it does not matter whether the doctor is human or robot. However, the hypothesis that Language and Frame play a role in *Medical Adherence* could not be confirmed.

Table 19. Bayesian ANOVA results for *Medical Adherence*

No.	Included effects	BF_{01}	BF_{10}
1.	Doctor	1.128	0.886
2.	Language	4.232	0.236
3.	Frame	5.335	0.187
4.	All main effects	25.364	0.039
5.	No. 4. + all 2-way interactions	795.698	0.001
6.	No. 5. + 3-way interaction	2481.664	4.03E-04

Expected Quality of Life. Table 20 shows partial output of the Bayesian ANOVA performed in JASP and "traditional steps" plus the step with the highest BF_{10} in the model-building sequence. First, Table 20 shows all main effects on their own, then all main effects included in one model. Subsequently, the model with the highest BF_{10} was included (All-main-effects plus the two interaction effects that include Doctor). Next, we added all 2-way interaction effects, and finally, we included the 3-way interaction effect. Table 20 shows both types of Bayes Factors.

We again hypothesized that there would be no main effect of Doctor, i.e. that the BF in favor of the null model compared to a model including the Doctor main effect would be high. If we look at the BF_{01} for the model including only the main effect of Doctor, we see that it is 6.257, indicating that there was evidence in favor of

the Null model. This is further supported by a low BF_{10}. Thus, it seems that who the doctor was (human or robot) did not matter, on its own, in predicting *Expected Quality of Life* scores.

However, this is not the best fitting model. The BF_{10} for the model including all main effects and two of the two-way interaction effects was highest of all tested models. This model was preferred to a model with only main effects by a Bayes Factor of 272.29, and to a model with all two-way interaction effects by a Bayes Factor of 3.02.

Thus, it seems that *Expected Quality of Life* scores were best explained by a combination of who the doctor was (human or robot), what language was used (affirmation or negation), and how the message was framed (positive or negative), and the interaction between Language and Doctor and Frame and Doctor. The interaction between Frame and Language did not seem to play a role. These results are partially in agreement with the frequentist results, the exception being that the main effect of Doctor was not statistically significant in the frequentist MANOVA. This is where the difference between our Bayesian ANOVA and the frequentist MANOVA becomes clear. Instead of testing each effect on its own, the Bayesian ANOVA tries to find the best fitting overall model for explaining *Expected Quality of Life* scores. In this context, the main effect of Doctor did result in an improved model fit, thus it was included even if its individual effect was non-existent (see $BF_{01} > 1$ for Doctor only model).

Thus, if we keep in mind the high BF_{01} for the Doctor only model, we can confirm our hypothesis that *Expected Quality of Life* scores were not affected by the type of doctor (human or robot). However, we could not confirm our hypothesis that a combination of affirmative language and positive framing led to higher *Expected Quality of Life* scores, as the preferred model did not include this interaction effect. Looking at Language on its own, we did not find evidence in favor of our hypothesis. Instead, we found that affirmation was associated with lower *Expected Quality of Life* scores ($M = 3.53$, $SE = .12$, 95% *CI* 3.29 – 3.78) than negation ($M = 4.57$, $SE = .12$, 95% *CI* 4.33 – 4.81). Furthermore, for Frame, we found that positive framing actually led to lower *Expected Quality of Life* scores ($M = 3.69$, $SE = .12$, 95% *CI* 3.45 – 3.93) than negative framing ($M = 4.41$, $SE = .12$, 95% *CI* 4.17 – 4.66). For a visual of the two interaction effects, please see Figure 1 and 2.

Table 20. Bayesian ANOVA results for *Expected Quality of Life*

No.	Included effects	BF_{01}	BF_{10}
1.	Doctor	6.257	0.16
2.	Language	2.58E-07	3.88E+06
3.	Frame	0.010	98.335
4.	All main effects	1.79E-08	5.58E+07
5a.	No 4. + Doctor*Language and Doctor*Frame	6.59E-11	1.52E+10
5b.	No. 4. + all 2-way interactions	1.99E-10	5.03E+09
6.	No. 5. + 3-way interaction	8.88E-10	1.13E+09

Effect of Ethics, Involvement, and Distance on outcomes: MANCOVA

For the participants that were in the Robot Doctor condition ($N = 134$), we also included questions about the doctor´s *Ethics*, the *Involvement* she stimulated, and affective *Distance* she provoked. To test the effect of these dependents on the outcome variables we treated them as covariates in a 2 (Language: Negation vs. Affirmation) x 2 (Framing: Negative vs. Positive) MANCOVA with all four outcome variables such as *Doctor Evaluation* and the like.

Multivariate Results. The multivariate results indicated that two of the covariates (*Ethics* and *Involvement*) had a significant effect together with the main effect of Language. The main effect of Language could be expected based on the results of the earlier MANOVA (Table 21).

Table 21. Multivariate results 2x2 MANCOVA

	Wilk's λ	*F*	*df hypo*	*df error*	*p*	Partial η^2
Ethics	0.902	3.38	4	124	.012	.098
Involvement	0.682	14.43	4	124	< .001	.318
Distance	0.945	1.79	4	124	.135	.055
Language	0.764	9.59	4	124	< .001	.236
Frame	0.992	0.26	4	124	.906	.008
Language * Frame	0.974	0.83	4	124	.506	.026

Univariate Results. The univariate results of the covariates suggested that that

Ethics had a univariate effect on *Doctor Evaluation* and *Message Evaluation*. In both cases, a higher *Ethics* score was related to a higher score on the outcome variable (Table 22).

The covariate *Involvement* had a significant effect on *Doctor Evaluation, Message Evaluation,* and *Medical Adherence*. Similar to *Ethics,* a higher score on *Involvement* was related to a higher score on the outcome variables.

Table 22. Univariate results 2x2 MANCOVA: Significant Multivariate Results

IV	DV	*df*	*F*	*p*	Partial η^2
Corrected Model	Doctor Evaluation	6	24.601	.000	.538
	Message Evaluation	6	6.223	.000	.227
	Expected Quality of Life	6	11.908	.000	.360
	Medical Adherence	6	6.697	.000	.240
Intercept	Doctor Evaluation	1	2.023	.157	.016
	Message Evaluation	1	2.060	.154	.016
	Expected Quality of Life	1	7.846	.006	.058
	Medical Adherence	1	17.531	.000	.121
Ethics	Doctor Evaluation	1	4.212	.042	.032
	Message Evaluation	1	9.428	.003	.069
	Expected Quality of Life	1	2.012	.159	.016
	Medical Adherence	1	.973	.326	.008
Involvement	Doctor Evaluation	1	51.036	.000	.287
	Message Evaluation	1	7.799	.006	.058
	Expected Quality of Life	1	.001	.981	.000
	Medical Adherence	1	12.636	.001	.090
Language	Doctor Evaluation	1	.001	.972	.000
	Message Evaluation	1	2.526	.114	.020
	Expected Quality of Life	1	37.761	.000	.229
	Medical Adherence	1	1.052	.307	.008
Error	Doctor Evaluation	127			
	Message Evaluation	127			
	Expected Quality of Life	127			

	Medical Adherence	127
Total	Doctor Evaluation	134
	Message Evaluation	134
	Expected Quality of Life	134
	Medical Adherence	134
Corrected Total	Doctor Evaluation	133
	Message Evaluation	133
	Expected Quality of Life	133
	Medical Adherence	133

Effect of conditions on Ethics, Involvement, and Distance: MANOVA

For the participants that were in the Robot Doctor condition ($N = 134$), we tested whether the scales *Ethics, Involvement,* and *Distance* were affected by the experimental conditions. Therefore, we performed a 2 (Language: Negation vs. Affirmation) x 2 (Frame: Negative vs. Positive) MANOVA on the three experiential variables.

Multivariate Results. As Table 23 shows, none of the multivariate effects were significant, indicating that the Language and Framing conditions did not affect *Ethics, Involvement,* and *Distance*.

Table 23. Multivariate results 2x2 MANOVA

	Wilk's λ	F	*df hypo*	*df error*	p	Partial η^2
Language	0.960	1.78	3	128	.155	.040
Framing	0.972	1.23	3	128	.302	.028
Language * Framing	0.980	0.86	3	128	.466	.020

Effect of Affordances and Use Intentions on outcomes: MANCOVA

For a subset of participants that were in the Robot Doctor condition ($N = 68$), two more experiential variables were measured, namely *Affordances* and *Use Intentions*. To test their effect on the outcome variables such as *Doctor Evaluation,* we included them as covariates in a 2 (Language: Negation vs. Affirmation) x 2 (Frame: Negative vs. Positive) MANCOVA with all four outcome variables.

Multivariate Results. The multivariate results show (Table 24) that both covariates had a significant effect together with the main effect of Language. Based on the results of our prior MANOVA, we should expect this main effect of Language.

Table 24. Multivariate results 2x2 MANCOVA

	Wilk's λ	*F*	*df hypo*	*df error*	*p*	Partial η^2
Affordances	.834	2.93	4	59	.028	.166
Use Intentions	.612	9.34	4	59	< .001	.388
Language	.444	18.49	4	59	< .001	.556
Frame	.952	0.75	4	59	.563	.048
Language * Frame	.986	0.21	4	59	.934	.014

Univariate Results. Table 25 shows that *Affordances* had a significant effect on *Message Evaluation* and *Medical Adherence.* In both cases, a higher *Affordances* score was related to a higher score on the outcome variable.

The covariate *Use Intentions* had a significant effect on *Doctor Evaluation* and *Message Evaluation.* As with *Affordances,* a higher score on *Involvement* was related to a higher score on the outcome variables.

Table 25. Univariate results 2x2 MANCOVA: Significant Multivariate Results

IV	DV	*df*	*F*	*p*	Partial η^2
Corrected Model	Doctor Evaluation	5	18.064	.000	.593
	Message Evaluation	5	10.982	.000	.470
	Expected Quality of Life	5	17.155	.000	.580
	Medical Adherence	5	10.262	.000	.453
Intercept	Doctor Evaluation	1	59.376	.000	.489
	Message Evaluation	1	29.082	.000	.319
	Expected Quality of Life	1	40.248	.000	.394
	Medical Adherence	1	76.232	.000	.551
Affordances	Doctor Evaluation	1	.008	.931	.000
	Message Evaluation	1	7.442	.008	.107
	Expected Quality of Life	1	.361	.550	.006
	Medical Adherence	1	8.091	.006	.115
UseIntention	Doctor Evaluation	1	32.645	.000	.345
	Message Evaluation	1	4.125	.047	.062
	Expected Quality of Life	1	2.645	.109	.041

	Medical Adherence	1	2.359	.130	.037
Language	Doctor Evaluation	1	.305	.582	.005
	Message Evaluation	1	3.905	.053	.059
	Expected Quality of Life	1	72.309	.000	.538
	Medical Adherence	1	.824	.368	.013
Error	Doctor Evaluation	62			
	Message Evaluation	62			
	Expected Quality of Life	62			
	Medical Adherence	62			
Total	Doctor Evaluation	68			
	Message Evaluation	68			
	Expected Quality of Life	68			
	Medical Adherence	68			
Corrected Total	Doctor Evaluation	67			
	Message Evaluation	67			
	Expected Quality of Life	67			
	Medical Adherence	67			

Effect of conditions on Affordances and Use Intentions: MANOVA

For the participants that were in the Robot Doctor condition with the extra two variables ($N = 68$), we tested whether *Affordances* and *Use Intentions* with regard to Robot Doctor were affected by the experimental conditions. We performed a 2 (Language: Negation vs. Affirmation) x 2 (Frame: Negative vs. Positive) MANOVA with both experiential variables.

Multivariate Results. As Table 26 shows, none of the multivariate effects were significant, indicating that the Language and Frame conditions did not affect *Affordances* and *Use Intentions.*

Table 26. Multivariate results 2x2 MANOVA

	Wilk's λ	*F*	*df hypo*	*df error*	*p*	Partial η^2
Language	0.951	1.06	3	62	.374	.049
Frame	0.994	0.14	3	62	.939	.006
Language * Frame	0.975	0.52	3	62	.668	.025

Bayesian Path Models

Analytic strategy

We used Mplus 7 to analyze the model displayed in Figure 8. To assess whether the model converged, we used the Gelman-Rubin criterion (Gelman et al., 2004). We applied a stricter cutoff value (bconvergence = .01) than the Mplus default value of .05. We also specified a minimum number of iterations using biterations = (10000). We ran two chains (chains = 2), and requested starting values based on the ML estimates (stvalues = ml). To further assess convergence, we inspected all trace plots. Finally, we chose the mean as our point-estimate of interest (other options are the median or the mode). Concerning priors, we used Mplus default priors, as there is no previous research that has studied this exact model, with these specific scales.

Results of model fit

This study measured five out of nine scales that originated from the theory of Interactively Perceiving and Experiencing Fictional Characters (I-PEFiC, Van Vugt, Hoorn, & Konijn, 2009). Therefore, we could only estimate a partial I-PEFiC model as depicted in Figure 8. This model included the encoding constructs *Ethics* and *Affordances*, and the responding constructs *Involvement*, *Distance*, and *Use Intentions.* We compared two versions of this model. The first model did not allow *Ethics* and *Affordances* to covary (in line with the I-PEFiC model), the second model, did allow this covariance (revised; represented by Figure 8).

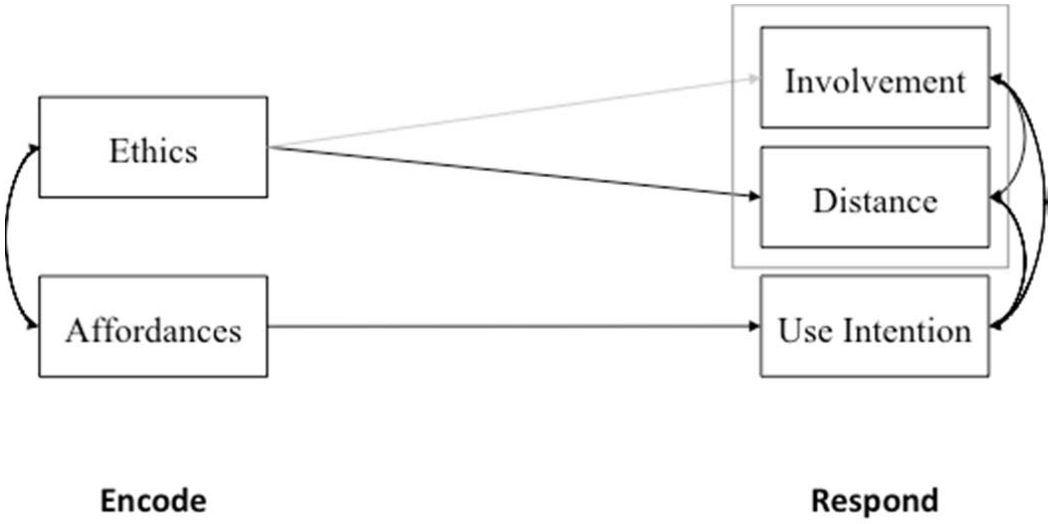

Figure 8. Partial I-PEFiC model (grey paths are not significant)

Looking at Table 27, the revised I-PEFiC model had both a smaller BIC value (Bayesian Information Criterion) and a smaller DIC value (Deviance) than the original I-PEFiC model, indicating that the original model fit the data worse than the revised model. Raftery (1995) stated that a BIC difference of >10 was strong evidence against the model with a higher BIC value (in this case, the original I-PEFiC model). Even though the BIC is slightly below 10, the DIC is higher than 10. Thus, we will continue with the revised model.

Table 27. BIC and DIC values for original and revised I-PEFiC models

	Original model	Revised model	Difference
BIC	1487.40	1478.91	8.49
DIC	1437.81	1426.13	11.68

Figure 8 shows that most paths are significant in a Bayesian sense (i.e. the 95% confidence interval excludes 0). There is one exception; *Ethics* has no effect on *Involvement*.

Table 28 shows the specific (standardized) parameter estimates of the model, as well as the R^2 of the dependent variables. *Ethics* is a negative predictor of *Distance*, such that higher evaluations of *Ethics* are related to lower feelings of *Distance*. *Affordances* are a positive predictor of *Use Intentions*, indicating that a positive opinion of the Robot Doctor's affordances are related to a higher evaluation of its *Use Intentions*.

Affordances and *Ethics* are positively correlated, meaning that the positive evaluations of the robot doctor's *Ethics* and *Affordances* often go together. *Distance* is negatively correlated with *Involvement* and *Use Intentions*, while *Involvement* and *Use Intentions* are positively correlated.

As for explained variance, the model explains 13.0% of the variance in *Distance*, 3.4% of the variance in *Involvement* (a non-significant predictor can still explain a limited amount of variance), and 21.3% of *Use Intentions*.

Table 28. Estimated Bayesian parameter coefficients of the revised I-PEFiC model

					95% CI	
	R^2	Stand. estimate	Estimate	Posterior *SD*	2.5%	97.5%
Distance on	0.130					

Ethics		-0.347	-0.423	0.121	-0.662	-0.193	*
Involvement on Ethics	0.034	0.149	0.133	0.098	-0.052	0.328	
Use Intentions on Affordances	0.213	0.434	0.449	0.162	0.154	0.784	*
Distance with							
Involvement		-0.589	-0.588	0.113	-0.837	-0.393	*
Use Intentions		-0.578	-0.741	0.241	-1.249	-0.300	*
Involvement with Use Intentions		0.673	0.672	0.211	0.280	1.103	*
Ethics Affordances		0.444	0.539	0.163	0.235	0.875	*

Note. * = Zero is not in the 95% CI

References

Burgers, C., Beukeboom, C. J., & Sparks, L. (2012). How the doc should (not) talk: When breaking bad news with negations influences patients' immediate responses and medical adherence intentions. *Patient Education and Counseling, 89*(2), 267-273. doi: http://dx.doi.org/10.1016/j.pec.2012.08.008

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). Boca Raton, FL, USA: Chapman & Hall/CRC.

Konijn, E. A., van de Schoot, R., Winter, S. D., & Ferguson, C. J. (2015). Possible solution to publication bias through Bayesian statistics, including proper null hypothesis testing. *Communication Methods and Measures, 9(4),* 280-302.

Love, J., Selker, R., Verhagen, J., Marsman, M., Gronau, Q. F., Jamil, T., Smira, M., Epskamp, S., Wild, A., Morey, R., Rouder, J., & Wagenmakers, E. J. (2015). *JASP (Version 0.6)* [Computer software].

Morey, R. D. & Rouder, J. N. (2015). *BayesFactor (Version 0.9.10-2)*[Computer software].

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25,* 111-163.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M., (2012) Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*, 356-374.

Van de Schoot, R., & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *European Health Psychologist, 16(2),* 75-84.

Van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & Aken, M. A. (2014). A gentle introduction to Bayesian analysis: applications to developmental research. *Child development, 85(3),* 842-860.

Van Vugt, H. C., Hoorn, J. F., & Konijn, E. A. (2009). Interactive engagement with embodied agents: An empirically validated framework. *Computer Animation and Virtual Worlds, 20,* 195-204. doi: 10.1002/cav.312

Stimulus Materials

(Dutch with English translation)

From Burgers et al. (2012).

Positive frame – Affirmation use

Goedemorgen, gaat u zitten. Ik zal er maar niet omheen draaien: Het nieuws dat ik u moet brengen **is goed**.

Zoals u weet, hebben we de afgelopen week allerlei tests gedaan om een diagnose te stellen voor uw klachten. We hebben een Röntgenfoto gemaakt en een test van Schober afgenomen. Uit al deze tests kwamen dezelfde resultaten. U bent positief getest voor de ziekte van Bechterew. Gezien de omstandigheden denk ik dat deze resultaten **goed zijn.**

Ik begrijp dat u vol met vragen zit op dit moment. Voor nu is het belangrijk te weten dat de ziekte van Bechterew een genetische ziekte is. De meeste patiënten vinden het **gemakkelijk** te leven met deze ziekte. Ik zal u specifieke medicatie voorschrijven. Met deze medicatie zal uw *levenskwaliteit* waarschijnlijk **vooruitgaan** de komende weken.

Ik raad u aan deze informatiefolder over de ziekte van Bechterew te lezen. Daarnaast wil ik graag een nieuwe afspraak maken voor over twee weken om de behandeling te evalueren.

Good morning, please sit down. I will not beat around the bush: The news I have to bring **is good**.

As you know, we have done a lot of tests in the past week to diagnose your complaints. We made an X-ray and took a Schober test. From all these tests the same results were obtained. You have been tested positively for Bekhterev's disease. In view of the circumstances, I think these results **are good**.

I understand you are full of questions right now. For now it is important to know that Bekhterev's disease is a genetic disease. Most patients find it **easy** to live with this disease. I will prescribe you specific medication. With this medication your quality of life will probably **progress** in the coming weeks.

I recommend reading this information leaflet about Bekhterev's disease. In addition, I would like to make a new appointment for about two weeks to evaluate the treatment.

Positive frame – Negation use

Goedemorgen, gaat u zitten. Ik zal er maar niet omheen draaien: Het nieuws dat ik u moet brengen is **niet slecht.**

Zoals u weet, hebben we de afgelopen week allerlei tests gedaan om een diagnose te stellen voor uw klachten. We hebben een Röntgenfoto gemaakt en een test van Schober afgenomen. Uit al deze tests kwamen dezelfde resultaten. U bent positief getest voor de ziekte van Bechterew. Gezien de omstandigheden zou ik stellen dat deze resultaten **niet slecht zijn.**

Ik begrijp dat u vol met vragen zit op dit moment. Voor nu is het belangrijk te weten dat de ziekte van Bechterew een genetische ziekte is. De meeste patiënten

vinden het **niet moeilijk** te leven met deze ziekte. Ik zal u specifieke medicatie voorschrijven. Met deze medicatie zal uw *levenskwaliteit* waarschijnlijk **niet achteruitgaan** de komende weken.

Ik raad u aan deze informatiefolder over de ziekte van Bechterew te lezen. Daarnaast wil ik graag een nieuwe afspraak maken voor over twee weken om de behandeling die ik u net voorgeschreven heb te beoordelen.

Good morning, please sit down. I will not beat around the bush: The news I have to bring is **not bad.**

As you know, we have done a lot of tests in the past week to diagnose your complaints. We made an X-ray and took a Schober test. From all these tests the same results were obtained. You have been tested positively for Bekhterev's disease. In view of the circumstances, I think these results **not bad.**

I understand you are full of questions right now. For now it is important to know that Bekhterev's disease is a genetic disease. Most patients find it **not hard** to live with this disease. I will prescribe you specific medication. With this medication your quality of life will probably **not deteriorate** in the coming weeks.

I recommend reading this information leaflet about Bekhterev's disease. In addition, I would like to make a new appointment for about two weeks to evaluate the treatment.

<u>Negative frame – Affirmation use</u>

Goedemorgen, gaat u zitten. Ik zal er maar niet omheen draaien: Het nieuws dat ik u moet brengen is **slecht.**

Zoals u weet, hebben we de afgelopen week allerlei tests gedaan om een diagnose te stellen voor uw klachten. We hebben een Röntgenfoto gemaakt en een test van Schober afgenomen. Uit al deze tests kwamen dezelfde resultaten. U bent positief getest voor de ziekte van Bechterew. Gezien de omstandigheden denk ik dat deze resultaten **slecht** zijn.

Ik begrijp dat u vol met vragen zit op dit moment. Voor nu is het belangrijk te weten dat de ziekte van Bechterew een genetische ziekte is. De meeste patiënten vinden het **moeilijk** te leven met deze ziekte. Ik zal u specifieke medicatie voorschrijven, echter zal met deze medicatie uw *levenskwaliteit* over de komende weken waarschijnlijk **achteruitgaan.**

Ik raad u aan deze informatiefolder over de ziekte van Bechterew te lezen. Daarnaast wil ik graag een nieuwe afspraak maken voor over twee weken om de behandeling die ik u net voorgeschreven heb te beoordelen.

Good morning, please sit down. I will not beat around the bush: The news I have to bring is **bad**.

As you know, we have done a lot of tests in the past week to diagnose your complaints. We made an X-ray and took a Schober test. From all these tests the same results were obtained. You have been tested positively for Bekhterev's disease. In view of the circumstances, I think these results are **bad**.

I understand you are full of questions right now. For now it is important to know that Bekhterev's disease is a genetic disease. Most patients find it **hard** to live with this disease. I will prescribe you specific medication. With this medication, however, your quality of life will probably **deteriorate** in the coming weeks.

I recommend reading this information leaflet about Bekhterev's disease. In addition, I would like to make a new appointment for about two weeks to evaluate the treatment.

Negative frame – Negation use

Goedemorgen, gaat u zitten. Ik zal er maar niet omheen draaien: Het nieuws dat ik u moet brengen is **niet goed.**

Zoals u weet, hebben we de afgelopen week allerlei tests gedaan om een diagnose te stellen voor uw klachten. We hebben een Röntgenfoto gemaakt en een test van Schober afgenomen. Uit al deze tests kwamen dezelfde resultaten. U bent positief getest voor de ziekte van Bechterew. Gezien de omstandigheden denk ik dat deze resultaten **niet goed** zijn.

Ik begrijp dat u vol met vragen zit op dit moment. Voor nu is het belangrijk te weten dat de ziekte van Bechterew een genetische ziekte is. De meeste patiënten vinden het **niet gemakkelijk** te leven met deze ziekte. Ik zal u specifieke medicatie voorschrijven, echter zal met deze medicatie uw *levenskwaliteit* over de komende weken waarschijnlijk **niet vooruitgaan.**

Ik raad u aan deze informatiefolder over de ziekte van Bechterew te lezen. Daarnaast wil ik graag een nieuwe afspraak maken voor over twee weken om de behandeling die ik u net voorgeschreven heb te beoordelen.

Good morning, please sit down. I will not beat around the bush: The news I have to bring is **not good**.

As you know, we have done a lot of tests in the past week to diagnose your complaints. We made an X-ray and took a Schober test. From all these tests the same results were obtained. You have been tested positively for Bekhterev's disease. In view of the circumstances, I think these results are **not good.**

I understand you are full of questions right now. For now it is important to know that Bekhterev's disease is a genetic disease. Most patients do **not** find it **easy** to live with this disease. I will prescribe you specific medication. With this medication, however, your quality of life will probably **not progess** in the coming weeks.

I recommend reading this information leaflet about Bekhterev's disease. In addition, I would like to make a new appointment for about two weeks to evaluate the treatment.

Questionnaire

(original version)

De algemene indruk die de robot-dokter op mij maakte was goed.

Volledig mee oneens ... ... Volledig mee eens

U krijgt nu een aantal vragen over deze robot-dokter.

Kunt u enkele symptomen noemen van de ziekte van Bechterew ?

Stijfheid van gewrichten.

Vermoeidheid.

Moedeloosheid.

Bent u bekend met de ziekte van Bechterew ?

Ja

Nee

Heeft u ervaring met dokter-patiënt gesprekken ?

Ja

Nee

Ondervond u problemen bij het bekijken van het filmpje ?

Ja

Nee

Beschrijf in 1 tot 3 zinnen waar de zojuist bekeken scene over ging.

Een nare medische boodschap.

Een negatief klinkende medische boodschap.

Een vervelende onbekende ziekte.

Wat is uw geslacht ?

Man

Vrouw

Nu volgen er een paar vragen over de boodschap die de robot-dokter bracht.

Het was begrijpelijk.

Volledig mee oneens ... ... Volledig mee eens

Was duidelijk.

Volledig mee oneens ... ... Volledig mee eens

Nam mijn hoop weg.

Volledig mee oneens ... ... Volledig mee eens

Was informatief.

Volledig mee oneens ... ... Volledig mee eens

Beleefd.

Volledig mee oneens ... ... Volledig mee eens

Meelevend.

Volledig mee oneens ... ... Volledig mee eens

Respectvol.

Volledig mee oneens ... ... Volledig mee eens

Tactvol.

Volledig mee oneens ... ... Volledig mee eens

Vriendelijk.

Volledig mee oneens ... ... Volledig mee eens

Wat is uw leeftijd in jaren ?

17

25

34

42

50
59
67
75
83
92
100
U kunt uw leeftijd schuiven met het pijltje.

De robot-dokter raadde u medicatie aan, wij willen u nu enkele vragen over deze medicatie stellen.

Het opvolgen van de adviezen van de dokter is altijd verstandig.
Volledig mee oneens ... ... Volledig mee eens

Het is een goed idee om de adviezen voor de behandeling op te volgen.
Volledig mee oneens ... ... Volledig mee eens

Wat is uw moedertaal ?
Nederlands
Anders

Wat is uw nationaliteit ?
Nederlands
Anders

Wat is uw hoogst afgeronde opleiding ?
Lager onderwijs
Middelbaar onderwijs
Hoger onderwijs / Wetenschappelijk onderwijs

Ik denk dat de robot-dokter mijn *levenskwaliteit* laag inschat.
Volledig mee oneens ... ... Volledig mee eens

De dokter sprak over de *kwaliteit van leven* met deze ziekte, wij willen u hier nu enkele vragen over stellen.

Op basis van dit gesprek schat ik mijn *levenskwaliteit* hoog in.
Volledig mee oneens ... ... Volledig mee eens
Ik schat de kwaliteit van de medicatie hoog in.
Volledig mee oneens ... ... Volledig mee eens

Ik denk dat de medicatie goed zal werken.
Volledig mee oneens ... ... Volledig mee eens

Ik denk dat de medicatie nuttig is.
Volledig mee oneens ... ... Volledig mee eens

I-PEFiC vragenlijst

Deze robot-dokter is incapabel.
Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter is kundig.
Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter is dom.
Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter is goedwillend.
Volledig mee oneens ... ... Volledig mee eens

Ik heb een goed gevoel bij de robot-dokter.
Volledig mee oneens ... ... Volledig mee eens

Ik voel me verbonden met de robot-dokter.
Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter is rechtvaardig.

Volledig mee oneens ... ... Volledig mee eens

Ik zou deze robot-dokter echt willen gebruiken.

Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter heeft goede bedoelingen.

Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter is gemeen.

Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter is handig.

Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter is aardig.

Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter heeft een naar karakter.

Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter is onhandig.

Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter is capabel.

Volledig mee oneens ... ... Volledig mee eens

De robot-dokter kwam koud over.

Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter is kwaadwillend.

Volledig mee oneens ... ... Volledig mee eens

De robot-dokter kwam warm over.

Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter is klunzig.

Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter is knullig.

Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter is vaardig.

Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter is onrechtvaardig.

Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter heeft slechte bedoelingen.

Volledig mee oneens ... ... Volledig mee eens

Ik zou me laten helpen door de robot-dokter.

Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter is vals.

Volledig mee oneens ... ... Volledig mee eens

Ik heb een slecht gevoel bij deze robot-dokter.

Volledig mee oneens ... ... Volledig mee eens

Ik voelde verwijdering tot de robot-dokter.

Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter is intelligent.

Volledig mee oneens ... ... Volledig mee eens

Ik had vriendschappelijke gevoelens voor de robot-dokter.

Volledig mee oneens ... ... Volledig mee eens

Ik zou de robot-dokter negeren.

Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter is onaardig.

Volledig mee oneens ... ... Volledig mee eens

Ik zou de robot-dokter best willen overslaan.

Volledig mee oneens ... ... Volledig mee eens

Ik zou de robot-dokter wegsturen.

Volledig mee oneens ... ... Volledig mee eens

Ik vind zo'n robot-dokter afstandelijk overkomen.

Volledig mee oneens ... ... Volledig mee eens

Ik voelde genegenheid voor de robot-dokter.

Volledig mee oneens ... ... Volledig mee eens

Ik zou de robot-dokter best nog een keer willen spreken.

Volledig mee oneens ... ... Volledig mee eens

Deze robot-dokter is betrouwbaar.

Volledig mee oneens ... ... Volledig mee eens

Hier onder volgt een laatste vraag.

Ik vind dat slecht-nieuws gesprekken door een robot gevoerd mogen worden.

Volledig mee oneens ... ... Volledig mee eens

.....

Dankjewel !