Multi-Perspective Persuasion by a Council of Virtual Coaches

Gerwin Huizing[1][0000-0001-9275-7470], Randy Klaassen[1][0000-0002-9296-6974], and Dirk Heylen[1][0000-0003-4288-3334]

[1] Human Media Interaction, Faculty of Electrical Engineering, Mathematics, and Computer Science, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
{g.h.huizing, r.klaassen, d.k.j.heylen}@utwente.nl

Abstract. Multi-perspective group persuasion by virtual characters has the potential to improve behaviour change support systems by making them more persuasive, satisfying to use, and effective at achieving value-added user outcomes. In this paper, we present a study into multi-perspective persuasion by a coaching team of virtual characters, in which we tried to investigate the effects of inter-coach discussion on multi-perspective persuasion on the topic of weight loss. We investigated the effects of inter-coach discussion during a coaching session with respect to the perception of the coaches, and their ability to coach. We compared two conditions. In one condition the coaches merely gave tips, and in the other they had brief discussions in between the tips. We used questionnaires, and held interviews to gain more insight on the perception of the council, and to determine whether commitment to tips changed due to inter-coach discussion. We found a minor difference in perception of the council between the conditions. We did not find perceived coaching ability to differ. Participants had a preference for the inter-coach discussion when they noticed the difference between conditions. There was a minor influence of inter-coach discussion on reflection on which approach to choose and why. There was a small increase in commitment to advice when inter-coach discussion had taken place. Finally, feedback from the interviews indicated the type of discussion the coaches have, influences how the participants perceived it. We conclude that inter-coach discussion between agents during group interaction, when noticed, is preferred by people. We also suggest that well-designed and pretested persuasive group discussion dialogues performed by virtual agents could have an effect on changing the opinions people have.

Keywords: Virtual agents, Multi-perspective persuasion, Group discussion.

1 Introduction

Xyrichis and Ream [15] characterise teamwork for healthcare settings as a dynamic process in which two or more healthcare professionals with backgrounds and skills that complement each other share the same health goals, and put in joint effort to asses, plan, or evaluate patient care. This works through interdependent collaboration, open communication, and shared decision-making. For those working in the team it leads to recognition of their individual contribution, increased motivation, and im-

proved mental health. They argue that for patients it leads to improved quality of care, increased value-added patient outcomes, and higher satisfaction with the provided services. Furthermore, teamwork improves task performance, learning, and communication, Effective teamwork requires team members to be open-minded and motivated [7].

In typical healthcare settings, a team consists of several professionals. Although the patient is traditionally not part of this team, shared decision-making with the patient is increasing in prominence in healthcare policy [5]. This is defined as an approach in which clinicians and patients share the best available evidence, and patients are supported to consider their options and achieve informed preferences [5]. Subsequent paragraphs, however, are indented.

Discussion can be seen as a vital part of open communication and shared decision-making, since the team members need to make sure they all share their expertise to come to the best solution with the patient. Previous work on group discussion has shown that it can change opinions, even on matters of fact [9]. It has also been shown in earlier work that group discussion improves cooperation when the group needs to solve some form of problem [11]. Though group discussion can change opinions, lead to new insights, and improve cooperation, there can be a negative side to it. For example, the conflict can interfere with effective decision-making, make group members frustrated and dissatisfied, and impede willingness of members to work together in the future [8].

The translation from real group interactions to group interactions with virtual agents has been studied in recent years by several groups. Multi-party negotiation was studied in previous work, in which participants were expected to negotiate as a US army captain with two virtual agents in a war zone setting. Each group member had their own goal in the negotiation [14]. In other work, researchers investigated the ability of agents in a multi-party interaction to increase human thinking and communication with the agents [4]. They did so by having a peer agent help participants guess the answers in a quiz given by a quiz master agent. The peer agent turned out to be both liked and effective in increasing responses. In more recent work, a study was done into the effects of persuasion done by multiple agents [10]. They found multiple agents to be more persuasive than a single agent. Though participants preferred user-directed persuasion, and felt it was more persuasive, vicarious persuasion seemed to be most effective at changing participants' behaviour.

Health issues are often multidisciplinary. In the real world a multidisciplinary healthcare team consisting of members all bringing their own perspectives works on health issues. A virtual healthcare team can use these different perspectives and present options to the user so they can make better informed decisions. In our virtual council of coaches setup we try to mimic a multi-disciplinary healthcare team. The virtual agents in the council could support the users in changing or adapting physical, emotional and/or mental behaviour. The virtual council consist of coaches with their own expertise. Users can interact with the team, and share their decision-making with them. The setup of the system supports the design of virtual coaches with different expertise, personality and coaching styles [1].

In this study, we explore teamwork in the form of group discussion, and multi-perspective persuasion focused on helping participants achieve a health goal (weight management). The fact that the system has multiple coaches discussing their opinion helps to show multiple perspectives on how to solve an issue, or answer a question. This could come with potential benefits. It could make the user reflect on their options, as they now get presented with different approaches to handle their issues, and multiple potential answers to their questions coming from several credible sources within one system. It might also make the system seem less biased as a whole, since it is presenting multiple perspectives on issues and questions to the user by multiple individuals, as opposed to a single individual. This could make it feel less like the opinions of one individual with their own biases talking to them, and instead like a talk with a group of individuals, each with their own biases, different perspectives, and ideas. These potential benefits do rely on the coaches each being seen as a separate individuals, and as having some credibility.

1.1 Aim of this paper

To explore the effects on people of group discussion and multi-perspective persuasion using a virtual council of coaches trying to help achieve a health goal (weight management), we aimed to find answers to the following question: "What is the effect of inter-coach discussion during a persuasive dialogue in a coaching session on the perception of the council of coaches, the perception of the council's ability to coach, and the council's actual coaching ability?"

To answer this question, we investigated the following research questions. Does inter-coach discussion during a persuasive group dialogue lead to a change:
1. in perception of a virtual council of coaches?
2. in perception of a virtual council of coaches' ability to coach?
3. in reflection on which approach to choose and why to choose it?
4. in commitment to follow a chosen approach?
5. in enjoyment of, and preference for an interaction?

2 Methods

2.1 Sampling and participants

We used the G*Power tool [6] to calculate a priori how large our sample size had to be able to detect a medium effect size (d = .50) or larger. With an error probability of .05 and a power of .80 we would require at least 34 participants. We decided to recruit our participants around the University of Twente, as we would have access to a large and diverse sample of people. The university attracts students and employees from all over the world, and the students and employees represent a fairly large age range. Furthermore, both students and employees at the University of Twente generally have a good understanding of the English language.

We recruited 45 participants at the University of Twente. All of them had the ability to work with a computer, and could converse effectively in English. Our sample

contained mostly students, as well as a few working adults. Due to technical issues disturbing our procedure during a few sessions, seven participants were excluded from analysis. The remaining group of 38 participants consisted of 22 male, and 16 female participants that were between 18 and 35 years old (M = 22.45, SD 3.438).

2.2   Materials

System The virtual council system was installed on the laptop that the researcher set up. The laptop was connected to an external screen, external speakers, and a computer mouse. The user interface consisted of an environment with a web browser in the background, a table at which a virtual council of three coaches sat down, and buttons that would appear for the participant to respond to the coaches. The buttons would only be on screen when the participant needed to respond to the coaches, and not while the coaches were talking.

The council of coaches consisted of three coaches that each had their own appearance, name, role, and expertise related to the topic of weight loss. Figure 1 shows the coaches in the scene. From left to right, it shows Harm (discussion lead, and mental coach), François (diet coach), and Alexa (physical activity coach).
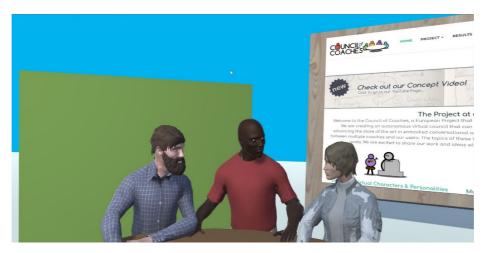

Fig. 1. Coaching scene from the perspective of the participant

Six tips were presented in two rounds. Each round consisted of three tips. Between the rounds Harm requested feedback from the participant on the tips. Each coach would give one tip per round using their expertise. The tips were offered in the following order:

Round 1.
1. François (diet): Lower your sugar intake.
2. Alexa (physical activity): Start a daily exercise routine consisting of jogging.

3. Harm (mental): Identify troubling thoughts, and tell yourself out loud to stop. Then try to introduce healthier thoughts.

Round 2.

4. Alexa (physical activity): Do strength training two to three times per week.

5. Harm (mental): Make sure you get enough sleep every night.

6. François (diet): Drink more water, especially shortly before each meal.

Questionnaires and interview We used a brief questionnaire asking for participants' age, gender, and experience interacting with virtual agents for demographic purposes. To answer our research questions, we used a part of the Godspeed questionnaire series [12], and an adjusted version of the Coaching Behaviour Scale for Sport (CBS-S) [3]. Within both of the questionnaires, the order of the questions was randomized for each condition for each participant. We chose for the Godspeed questionnaire series, because it is a well-known and often used set of questionnaires in the virtual agent community that measures the perception participants have of virtual agents. We chose for an adjusted version of the CBS-S since it contains many items that measure the perception participants had of the coaching ability of their coach. We applied these items to have participants evaluate the virtual coaching team. Furthermore, we used interview questions developed to get more in depth answers from the participants.

For the Godspeed questionnaire series, we selected the anthropomorphism, animacy, likeability, and perceived intelligence questionnaires. We left out the perceived safety questionnaire, as we did not expect our interactions to have a strong impact on participants with regards to anxiety, agitation, or surprise.

The original CBS-S we found in [3] contained several items that were not relevant to the interactions in our experiment. For that reason, we did not use items on the scales of physical training and fitness, technical skills, competition strategies, personal rapport, and negative personal rapport (i.e. items 1 to 15, and items 27 to 47). On the scale of mental preparation, we did not use the item regarding performance under pressure (i.e. item 16), as the coaching interactions were about weight management, and did not address performing under pressure. On the scale of goal settings, we did not use the items regarding monitoring of progress, identifying target dates for attaining goals, and setting long-term goals (i.e. items 22, 24, and 25), as the coaching interactions were not about progress, planning, or setting long-term goals. The coaching interactions were focused more on how to behave in the short term, and weight management tips. For all the items that we used, we rephrased them from "the coach(es) most responsible for my" to "my coaching team", as we used a virtual council of coaches. Furthermore, several items relevant to the interactions in our experiment were added under a new "coaching quality" scale. These were the following items:

1. My coaching team helps me to be motivated and inspired by others.

2. My coaching team helps me to discover which things help me to attain and maintain my healthy weight better.

3. My coaching team had the right knowledge and abilities to give good coaching.

4. My coaching team gives advice of good quality.

The interview questions were about the experience of working with the system, behaviour of the coaching team and interactions with them, advice chosen by partici-

pants, commitment to the advice, and reasoning for the commitment, intention to use the system again, and recommendation of the system.

2.3   Experimental design

The experiment used a 1 × 2 within-subjects counterbalanced measures design. The independent variables were the following two interaction conditions:

Condition 1: The three coaches each presented one tip, and some explanation on their own tip. Then coach Harm asked how the participant liked the advice so far. Then the coaches presented another tip each, again with some explanation on their own tip. At the end, coach Harm asked the participant how they liked all the advice, and which of the six tips presented to them during the interaction the participant preferred. No additional information was given about the tips for the tip preference question by Harm. The participant made their choice using what they remembered of the explanation given about the tips earlier in the interaction by the coaches. Once the participant indicated which advice they liked best, the coaches would offer words of encouragement to the participant, wish the participant luck, and close out the conversation. In Condition 1, the coaches did not interact with each other between giving advice, and simply took the turn from the previous coach to start presenting their own advice.

Condition 2: The three coaches presented the same tips in the same order, and with the same explanation as in Condition 1. Harm asked the same questions of the participant between the rounds, and after the second round, as he did in Condition 1. No additional information was given about the tips for the tip preference question by Harm. The participant made their choice using what they remembered of the explanation given about the tips earlier in the interaction by the coaches. The coaches then offered the same words of encouragement, and closed out the conversation in the same way. In contrast to Condition 1, when transitioning to the next advice, the coaches would briefly interact with each other regarding their advice, mimicking a real-life group discussion. These interactions consisted of the coach that would present their advice next remarking briefly on their thoughts about the previous tip given to the user. Then they would mention the importance of their own upcoming advice. The coach giving the previous tip would respond mildly critically to this, and then asked them to elaborate. The coach presenting their advice next would then start to present that advice, with the same content as in Condition 1. The discussions between each advice lasted roughly twenty to thirty seconds each.

2.4   Procedure

Participants were individually tested. Each of them was let into the experiment room with the setup being ready. The experiment was conducted in English. The participants were asked to read the information letter and sign the informed consent form. Afterwards, the researcher would sign the informed consent form, and would offer a copy of the information letter and informed consent form. Then, the researcher would explain the procedure and tasks to the participant.

After the explanation, the participant would fill out their demographic information on a tablet. Then the researcher would explain that the introduction had the purpose of getting to know the coaching team, and learning how to work with the interface. They could ask the researcher questions. In the introduction the coaches gave their name, and briefly explained their expertise.

The participant then had two interactions with the coaches, specified in the Experimental Design section (conditions). After each interaction, they answered the Godspeed questionnaires [12], followed by the adjusted CBS-S [3] on a tablet.

Once the participant was done with the two interactions and rounds of questionnaires, the researcher verbally asked for permission to record the interview. If consent was given, they proceeded to conduct an interview with the participant. The topics discussed are described in the Research materials section.

3 Results

3.1 Quantitative measures

Construct Reliability The constructs of anthropomorphism, animacy, likeability, and perceived intelligence of the Godspeed questionnaires have been found to have good internal consistency (all Cronbach's alphas > .70) in earlier studies, as described in [2]. These constructs also had good internal consistency in both conditions in our experiment (see Table 1).

In an earlier study, researchers found that the constructs of the CBS-S had good internal consistency [3]. They reported alpha coefficients of all scales above .85 (N=205). Our scales generally had satisfactory internal consistency of constructs for both conditions (see Table 1). The exceptions were in Condition 2 on our mental preparation scale and goal setting scale. Presumably, the slightly low alpha coefficients were due to the low amount of items and participants.

Table 1. Construct Reliability Statistics (N = 38)

Condition	Questionnaire	Construct	Alpha
1	Godspeed Questionnaires	Anthropomorphism	.86
1	Godspeed Questionnaires	Animacy	.83
1	Godspeed Questionnaires	Likeability	.87
1	Godspeed Questionnaires	Perceived Intelligence	.72
1	Adjusted CBS-S	Mental Preparation	.76
1	Adjusted CBS-S	Goal Setting	.71
1	Adjusted CBS-S	Coaching Quality	.70
2	Godspeed Questionnaires	Anthropomorphism	.81
2	Godspeed Questionnaires	Animacy	.83
2	Godspeed Questionnaires	Likeability	.89
2	Godspeed Questionnaires	Perceived Intelligence	.80
2	Adjusted CBS-S	Mental Preparation	.68

| 2 | Adjusted CBS-S | Goal Setting | .68 |
| 2 | Adjusted CBS-S | Coaching Quality | .79 |

Godspeed questionnaires Difference scores were calculated by subtracting scores in Condition 1 from Condition 2 on the Godspeed questionnaires. The difference score for the anthropomorphism questionnaire (M = -.04, SE = 0.1), indicating lower reported anthropomorphism in Condition 2, was not significantly different from 0, t(37) = -.383, p = .704; it represented a small effect size of r = .06. The difference score for the animacy questionnaire (M = -.15, SE = .08), indicating lower reported animacy in Condition 2, was not significantly different from 0, t(37) = -1.928, p = .062; however it represented a medium effect size of r = .30, and showed a trend towards significance. The difference score for the likeability questionnaire (M = -.15, SE = .12), indicating lower reported likeability in Condition 2, was not significantly different from 0, t(37) = -1.310, p = .198; it represented a small effect size of r = .21. The difference score for the perceived intelligence questionnaire (M = -.1, SE = .1), indicating lower reported perceived intelligence in Condition 2, was not significantly different from 0, t(37) = -1.024, p = .313; it represented a small effect size of r = .17.

Adjusted CBS-S Difference scores were calculated by subtracting scores in Condition 1 from Condition 2 on the adjusted CBS-S scales. The difference score for the mental preparation scale (M = -.03, SE = .12), indicating lower reported coaching related to mental preparation in Condition 2, was not significantly different from 0, t(37) = -.285, p = .778; it represented a small effect size of r = .05. The difference score for the goal setting scale (M = .00, SE = .09), indicating no difference in reported coaching related to goal setting between Condition 1 and Condition 2, was not significantly different from 0, t(37) = .000, p = 1.000; it represented a minuscule effect size of r < .00. The difference score for the coaching quality scale (M = .12, SE = .09), indicating higher reported coaching quality in Condition 2, was not significantly different from 0, t(37) = 1.321, p = .195; it represented a small effect size of r = .21.

3.2 Qualitative measures

In this section, we briefly describe the results of the interviews with our sample of 38 participants. Remarks were on the interaction participants had with the coaches using the interface, noticed condition differences and preferences, chosen advice, reasoning for this choice, and commitment to following the advice.

Interface and interaction Part of the participants indicated they liked the interaction (10). Participants remarked that it was not as good to talk to the coaches as to real professionals, but it still felt quite nice, and natural (5). On the other hand, participants mentioned that their ways to respond to the coaches felt limited (15), and indicated that they could not always voice their opinions, and thoughts during the interac-

tion (7). Furthermore, participants remarked the coaches behaved in a scripted and one-sided way (18), and felt like their input did not matter much, as they would get the same reply regardless (8). The behaviour of the coaches was described as being robotic and unnatural. For example, the movement of the coaches felt stiff at times, the communication seemed to lack a natural flow, and the response time was slow at times (9). Participants also mentioned that the coaches were talking more to each other than to them (3), and the coaches did not clearly signal to the participant when to respond (2).

Condition differences and preferences Participants were all asked to indicate whether they noticed a difference between the two interaction conditions, and if so, what they thought it was. They were also asked about their preference between the conditions. For an overview of the identified condition differences and condition preferences for each condition order, see Table 2. We will detail the results in the rest of this section.

Table 2. Identified differences Condition 1 and Condition 2, and preferences (N = 38)

Order	Identified difference	Preference Condition 1	Preference Condition 2	No Preference
1-2	Correctly identified	0	9	2
1-2	Incorrectly identified	2	0	0
1-2	Not identified	0	1	5
2-1	Correctly identified	1	1	1
2-1	Incorrectly identified	6	0	1
2-1	Not identified	0	0	9

Part of the participants that started with Condition 1 indicated they noticed a difference (13). Part of them correctly identified the main manipulated difference (11), such as remarking it felt like more of a discussion in Condition 2. The majority of them preferred Condition 2 for (9), and the minority had no strong preference (2). Several participants felt they noticed differences, but these experienced differences were not present (2), such as feeling Condition 2 gave less options. These participants preferred Condition 1 (2). The remaining participants indicated they could not find any difference between the conditions (6). The minority preferred Condition 2 (1), and the majority had no preference (5).

Part of the participants that started with Condition 2 indicated they noticed a difference (10). Some of them correctly identified the main manipulated difference (3), such as saying that in Condition 2 the coaches had aggressive discussions with each other that were not present in Condition 1. One of them preferred Condition 1 (1), one Condition 2 (1), and one had no strong preference (1). Several participants felt they noticed differences, but these experienced differences were not present (7), such as Condition 1 feeling more smooth, and the coaches addressing each other more in Condition 1. The majority preferred Condition 1 (6), and the minority had no strong

preference (1). The remaining participants indicated they could not find any difference between the two conditions (9). None of them had a strong preference (9).

When looking at this data (see Table 2), what stands out is the higher amount of participants that started with Condition 1 correctly identifying the manipulated difference, as compared to those that started with Condition 2. Furthermore, we see that participants that did correctly identify the main manipulated difference generally preferred Condition 2, those that incorrectly identified this difference generally preferred Condition 1, and those that could not identify any differences at all generally had no preference.

Advice choice, reasoning and commitment Participants were asked to indicate which advice they chose in each condition. They gave an explanation on why they chose that advice, and rated their commitment from one to seven to try it during the next month.

The majority of participants that started with Condition 1 chose the same advice in both conditions (14). The other participants decided on different advice (5). The main reasons mentioned were the novelty of the information (8), advice serving as a reminder of importance to them (6), recommended behaviour being easy to perform (11), advice having the most impact on their life as a whole (5), advice best matching their needs and goals (4), recommended behaviour already being performed (4), and quality of the interaction in Condition 2 convincing them (3). Participants averagely rated their commitment to the advice chosen at 5.26 in Condition 1, and at 5.53 in Condition 2 on a seven point scale.

The majority of participants that started with Condition 2 chose the same advice in both conditions (14). The other participants decided on different advice (5). The main reasons mentioned were the advice being important and often forgotten (7), importance of the advice to them (7), behaviour in the advice being something they were committed to (6), advice most applied to them (12), recommended behaviour being easy to perform (13), novelty of the information (8), advice serving as a reminder of importance to them (5), recommended behaviour already being performed (5), and advice having the most impact on their life as a whole (3). Participants averagely rated their commitment to the advice chosen at 5.55 in Condition 1, and at 5.71 in Condition 2 on a seven point scale.

4   Discussion

4.1   Research question 1: Perception of council of coaches

Our statistical analysis showed no significant effect on any of the used Godspeed questionnaires. We saw a trend towards significance for a more positive rating of Condition 1 on the animacy questionnaire (M = -.15, SE = .08, t(37) = -1.928, p = .062), with a medium effect size (r = .30). The higher animacy in Condition 1 could be related to participants mentioning during interviews that they noticed differences

between the conditions, such as the coaches moving and speaking more fluidly in Condition 1 as compared to Condition 2.

4.2 Research question 2: Perception of council of coaches' ability

Our statistical analysis showed no significant effect on any of the adjusted CBS-S scales. We found an insignificant small effect indicating a higher rating on the mental preparation scale in Condition 1 ($M = -.03$, $SE = .12$, $t(37) = -.285$, $p = .778$, $r = .05$), an insignificant minuscule effect indicating no difference between Condition 1 and Condition 2 on the goal setting scale ($M = .00$, $SE = .09$, $t(37) = .000$, $p = 1.000$, $r < .00$), and an insignificant small effect indicating a higher rating on the coaching quality scale in Condition 2 ($M = .12$, $SE = .09$, $t(37) = 1.321$, $p = .195$, $r = .21$). To support the finding of the higher rating on the coaching quality scale in Condition 2, we can look at the participants stating that they picked advice in Condition 2, because the quality of the interaction was better there (3). This leads us to believe that there could be a small effect on perceived coaching ability due to inter-coach discussion. It may have been masked here by a substantial amount of participants not noticing the inter-coach discussion.

4.3 Research question 3: Reflection on choices that were made

During the interviews, participants often indicated they made a choice based on personal reasons, such as being reminded of the importance of the chosen advice (11), or the novelty of the information (16). As previously mentioned, some participants did mention picking advice in Condition 2, because the quality of the interaction was better than in Condition 1 (3). This did influence their choice, according to them. Considering the amount of times reasons were mentioned, some participants mentioning the better quality of interaction in Condition 2 does suggest there was an impact of the inter-coach discussion on the reflection people had about what approach and advice to choose, and why to choose it, but only a small one.

4.4 Research question 4: Commitment to a chosen approach

In the interviews, participants indicated a stronger commitment to their chosen advice in Condition 2, as compared to Condition 1. This was the case for those that started with Condition 1 (Condition 1: $M = 5.26$, Condition 2: $M = 5.53$), and those that started with Condition 2 (Condition 1: $M = 5.55$, Condition 2: $M = 5.71$). Though the differences were not huge, and many participants gave similar ratings in both conditions, these differences do indicate that the inter-coach discussion increased reported commitment by the participants.

4.5   Research question 5: Interaction preference

During the interviews, participants were asked to indicate the differences they perceived between the conditions, and which condition they preferred. We saw that 15 of the participants could not identify any difference, and 9 participants identified a difference that was not present. The remaining 14 participants did identify the main manipulated difference. Those that could not identify the difference generally did not have a strong preference (14 of 15 participants), and those that incorrectly identified the difference generally had a preference for Condition 1 (8 of 9 participants). Finally, those that did identify the main manipulated difference generally preferred Condition 2 (10 of 14 participants). This indicates a potential change in preference for an interaction due to inter-coach discussion. Which direction this change goes seems to be linked to whether the participant perceived the inter-coach discussion (preference inter-coach discussion), thought they perceived another difference which was not there (preference no inter-coach discussion), or did not notice any difference (no preference).

5   Conclusion

In this study, we set out to investigate the effect of inter-coach discussion during a persuasive coaching dialogue on the perception of a virtual council of coaches, and their ability to coach. Firstly, the one effect inter-coach discussion had on the perception of the council, was a marginally significant decrease in perceived animacy when inter-coach discussion took place. This was backed up by remarks by participants in the interviews. As our setup used text-to-speech voices, and had limited movement by the virtual agents, we believe the longer exposure to them in Condition 2 could explain the reported lower animacy, as well as the related remarks. However, further study would be necessary.

Furthermore, the inter-coach discussion led to a change in preference for interactions. When participants noticed the inter-coach discussion, they generally preferred it. Those who noticed the discussion, may have preferred it for fleshing out the conversation and coaches, and feeling more similar to a normal talk with people. Furthermore, the perceived shared decision-making may have made them more motivated, and satisfied with the outcomes [15].

Additionally, we found reflection on which approach to choose and why was influenced by inter-coach discussion. The effect was quite minor. When looking at the interview data, we interpreted the majority of reasoning to not be motivated by the differences between the interactions. This may have had to do with the fact that much of the given advice consisted of quite general, well-known ways to lose weight. Furthermore, the length of the interactions with our setup, that used text-to-speech voices and agents with limited movement, might have reduced the impact of the difference between conditions. In future work on multi-party inter-agent discussion, more time should be spent on the designing and pretesting of dialogues. In making well-designed dialogues, the focus should be on improvements in content, length, strategies employed by the agents, and interpersonal behaviour between the agents and towards the

user. Many other factors could also be considered, based on the study the dialogues would be used in.

We also found a small, consistent increase in reported commitment to follow a chosen approach when inter-coach discussion took place compared to when it did not. Especially interesting is the fact that participants starting with Condition 2 felt this way. A large majority of these participants could not identify any differences between the conditions, or incorrectly identified the difference, and had no preference for any condition, or preferred Condition 1. Yet, on average, they still felt more committed to the advice chosen in Condition 2. This seems to indicate that whether or not the inter-coach discussion was noticed or part of the preferred interaction, it was effective in improving reported commitment to an approach. This points to potential increased motivation, said to occur for those working in a team [15]. An alternative explanation would be vicarious persuasion mechanisms being at work, with the inter-coach discussion having more power to persuade [10]. The increase in commitment, as well as potential causes for this increase should be investigated in future work.

Finally, our results suggest that group discussion with inter-agent interaction can be a valuable addition to multi-party interaction with virtual agents, if the group discussion is noticed. Among other things, it has shown support for the notion that agents can have an effect in changing opinions through group discussions, similar to humans [9]. A study with well-designed and pretested group discussion dialogues could help to further investigate the findings of this study, as well as look into the causes. We plan to conduct such a study in our future work, and delve further into the design, effects, and efficacy of group discussion with virtual agents. Since the system used for this study enables us to design multi-party coaching teams, in future work we could also look into incorporating design principles, such as the Persuasive System Design model [13]. For example, the group interaction between the agents taking place could make the interaction feel more realistic and social (social support, dialogue support, credibility support).

Our study had several limitations. Firstly, we used a convenience sample for this study. This makes it hard to say how representative the results are. Additionally, the interactions we presented were lengthy. This may have led to distorted results, as participants could have had trouble maintaining their focus. In future studies, the interaction should be kept short, to make maintaining focus easier. Furthermore, our use of a within-subject design may have led to the issue of negative carryover effects due to fatigue and trouble staying focused in some participants. Since our design was counterbalanced by presenting the two conditions in a random order for each participant, we were able to mitigate the effect of this issue on our data to some extent. Offering breaks between the interactions in future studies to help prevent fatigue and lack of focus could further reduce the impact of this issue. Moreover, the lack of body movement by the agents and the text-to-speech voices might have also contributed to a lack of focus by participants. We received feedback from participants indicating this might have been the case. Another limitation was the substantial amount of participants that did not notice the inter-coach discussion. This made it harder to discern differences in perception of said discussion among users. This may have been caused by the aforementioned fatigue and trouble staying focused in some participants. An-

other cause could be that the discussion was not noticeable enough, as to participants it may have been a relatively small change in a larger interaction. As the dialogue was not tested before being used, it could have also been a dialogue design issue. The order in which the conditions were presented might also be part of the problem, as those who went through Condition 2 first did not notice it nearly as often. The disappearance of something relatively small in the interaction that is the same in every other way might not stand out as much compared the addition of something new when participants are already having trouble maintaining their concentration. Lastly, several participants remarked that the discussion seemed aggressive and competitive. This would be in line with [7] mentioning the importance of open-mindedness for effective teamwork, as well as the mention by [8] of group conflict having the potential to lead to frustration and dissatisfaction by group members. It would be of interest to delve into what form of inter-coach discussion is preferred, and why, as this could improve the effects of the discussion.

As virtual agents move into the realm of coaching, and develop the ability to manage group interactions, the opportunity to show multiple perspectives, and leverage group discussion presents itself. We hope this study contributes to the growing body of work on group interaction, multi-perspective persuasion, and group discussion performed with virtual agents.

References

1. op den Akker, H., op den Akker, R., Beinema, T., Banos, O., Heylen, D., Bedsted, B., Pease, A., Pelachaud, C., Traver Salcedo, V., Kyriazakos, S. and Hermens, H. Council of Coaches - A Novel Holistic Behavior Change Coaching Approach. In Proceedings of the 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health - Volume 1: ICT4AWE, ISBN 978-989-758-299-8, pages 219-226. DOI: 10.5220/0006787702190226 (2018).
2. Bartneck, C., Croft, E., and Kulic, D.: Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Social Robotics **1**(1), 71-81 (2009).
3. Côté, J., Yardley, J., Hay, J., Sedgwick, W., and Baker, J.: An exploratory examination of the coaching behavior scale for sport. Avante Research Note **5**(2), 82-92 (1999).
4. Dohsaka, K., Asai, R., Higashinaka, R., Minami, Y., and Maeda, E.: Effects of conversational agents on human communication in thought-evoking multi-party dialogues. Proceedings of the {SIGDIAL} 2009 Conference. Association for Computational Linguistics, London, UK, 217-224 (2009).
5. Elwyn, G., Frosch, D., Thomson, R., Joseph-Williams, N., Lloyd, A., Kinnersley, P., Cording, E., Tomson, D., Dodd, C., Rollnick, S., Edwards, A., and Barry, M.: Shared deci-

sion making: a model for clinical practice. Journal of General Internal Medicine **27**(10), 1361-1367 (2012).

6. Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A: G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, **39**, 175-191 (2007).

7. Ingram, H., and Desombre, T.: Teamwork in health care. Journal of Management in Medicine **13**(1), 51-59 (1999).

8. Jehn, K., Greer, L., Levine, S., and Szulanski, G.: The effects of conflict types, dimensions, and emergent states on group outcomes. Group Decision and Negotiation **17**(6), 465-495 (2008).

9. Jenness, A.: The role of discussion in changing opinion regarding a matter of fact. The Journal of Abnormal and Social Psychology **27**(3), 279-296 (1932).

10. Kantharaju, R., Pease, D., De Franco, D, and Pelachaud, C.: Is two beter than one? Efects of multiple agents on user persuasion. Proceedings of the 18th International Conference on Intelligent Virtual Agents, IVA 2018. CoRR abs/1904.05248 (2019).

11. Meleady, R., Hopthrow, T., and Crisp, R.: The group discussion effect: integrative processes and suggestions for implementation. Personality and social psychology review: an official journal of the Society for Personality and Social Psychology **17**, (2012).

12. Official Godspeed Questionnaire Series web page, http://www.bartneck.de/2008/03/11/the-godspeed-questionnaire-series/. Last accessed 17 Nov 2019.

13. Oinas-Kukkonen, H., and Harjumaa, M.: Persuasive systems design: key issues, process model, and system features. Communications of the Association for Information Systems **24**, 485-500 (2009).

14. Traum, D., Marsella, S., Gratch, J., Lee, J., and Hartholt, A.: Multi-party, multi-issue, multi-strategy negotiation for multi-modal virtual agents. Intelligent Virtual Agents. IVA 2008. Lecture Notes in Computer Science, vol 5208. Springer, Berlin, Heidelberg (2008).

15. Xyrichis, A., and Ream, E.: Teamwork: a concept analysis. Journal of Advanced Nursing **61**(2), 232-241 (2008).