Evaluating Semantic Co-creation by Using a Marker as a Linguistic Constraint Tool in Shared Cognitive Representation Models

Stefan Schneider and **Andreas Nürnberger**

1 Introduction

Information overload has become one of the most critical challenges in humans history. It has been shown that speech, writing, math, science, computing and the Internet are based on independent languages, which together form an evolutionary chain of languages as response to information overload (Logan 2006). Recent technological developments like internet of things (Färber et al. 2020) or cognitive augmented reality (Chi 2009) make clear that this chain continues to advance. Researchers must therefore ask themselves which approaches are suitable to cope with the new levels of complexity.

Cognitive representation models are a key element of the presented evolution, not only with the individual, but also where they can be used as artefacts in conversation. A cognitive representation model can be understood as an abstract model, from which an individual can infer the relationship of objects to one another his environment (Kaplan et al. 2017). Typically, the objects are related based on their properties. Considering e.g. a scale from "tiny" to "big", a "needle" would very strongly relate to "tiny", whereas "mountain" relates more to the "big" property. Using cognitive representation models in a collaborative manner can improve situations in which someone tackles a problem solving task, such as human-robot-interaction (Spranger 2016), within a complex environment. In such situations, people communicate successfully if they come to the conclusion that they are talking about the same things and their cognitive representations converge (Brennan 2005). Such a convergence is known as semantic co-creation (Gergen 2009). Evaluating this phenomenon becomes dif-

---

S. Schneider (✉) · A. Nürnberger
Data and Knowledge Engineering Group, Faculty of Computer Science,
Otto-von-Guericke-University Magdeburg, 39106 Magdeburg, Germany
e-mail: stefan.schneider@sschneider.de
URL: http://www.dke-research.de

A. Nürnberger
e-mail: andreas.nuernberger@ovgu.de

ficult under realistic conditions, while many factors (like technical communication problems) can bias the given observation (Kraut et al. 2002). Simple collaborative identification tasks (also named referring expression tasks) enable the evaluation of shared cognitive representation models under controlled laboratory settings and allow to observe the progress on semantic co-creation moment-by-moment (Brennan 2005). Findings on how to reach a state of semantic co-creation more easily are helpful in developing adaptive systems that make the complexities of the environment easier to use.

Previous work on referring expression tasks have got a tradition in evaluating collaborations which have a shared space (Kraut et al. 2002; Brennan et al. 2008; Neider et al. 2010; Müller et al. 2013; Hanrieder 2017) as well as a shared cognitive representation model (Brennan 2005; Keilmann et al. 2017). A basic example could be two people who try to find a particular street of a city together by sharing a geographical map. One person who is familiar with the location of the street could explain the route to this target by referring to places in relation to the target street which both participants are familiar with.

In one specific case researchers wondered about the benefit of using markers within these shared artefacts to improve the coordination behaviour based on visual evidence. The question arises if a shared marker can support the participants to achieve a state of semantic co-creation based on a shared cognitive representation model. A shared marker can be anything (like shared gaze (Brennan et al. 2008), shared mouse (Müller et al. 2013) or shared location (Keilmann et al. 2017)), which can be used in shared space or cognitive representation model as a spatial indicator (Müller et al. 2013). Results in this field state that shared markers are in general a beneficial tool (Brennan 2005; Brennan et al. 2008; Hanna and Brennan 2007; Neider et al. 2010; Müller et al. 2013). This becomes obvious when we reconsider the example about human behaviour regarding travelling. In the simplest case, a marker could be a finger of a participant moving across the map in order to explain or support a description non-verbally. If participant A says: "Drive straight through the small street until the next crossing is coming!" Participant B moves his finger along the road, in a manner he has understood the utterance of participant A. Once participant B has moved in a sufficient manner participant A will continue his description, e.g. "Ok! From this crossing, then turn left again." In a case of misconception, participant A would e.g. say to participant B: "No! I meant another street." The finger as some kind of marker applies the participant's given conception onto a map, which indicates to the other participant which aspects were comprehended correctly. Using such a marker (pointing to portions of a map) in addition to a cognitive representation model (a map) appears to be successful in solving a collaborative language task. All participants are informed promptly, using the model, about what has currently been understood (Kraut et al. 2002).

Despite the obvious benefit of using a shared marker, the current research results are not as clear as it might be expected. Specially, the problem is that even any study enforces the usage of a shared marker, while the task durations are very short. For example, Brennan observed task durations between 10 and 20 s (Brennan 2005). Such short durations let us infer that no real team interaction occurred, and then

using a marker is no benefit but only a requirement to finish the task. Based on this contribution, we want to assess the benefit of a shared marker when it is optionally used in comparison to an increased decision complexity. Decision complexity is a user constructed criterion based on the number of alternatives available (Payne 1976). It has been shown, that configuring structural properties of a shared space (Kraut et al. 2002) or cognitive representation model (Keilmann et al. 2017) can influence the perception and even the communicative success. If we use a shared marker optionally, then it can be understood in the manner of a linguistic constraint tool. Such tools relate in some way to linguistic constraints, which cover by symbols the application of dynamic constraints in language use. We hypothesize that the value curve of a linguistic constraint tool based on a given cognitive complexity determines if it is useful or not in given situation of collaboration. In this study we will show that if a shared marker becomes optional under less decision complexity, then it becomes too expensive to use them.

This contribution is structured in the following manner: Collaborative task settings using conversation can be explained by using the contribution model (Sect. 2). Based on the contribution model, semantic co-creation happens when the grounding criterion is reached. There are forces that influence the nature of contributions within the discourse, named linguistic influences. Linguistic influences on the grounding criterion have yet to be investigated in research. Hence, we explain in detail the concept of linguistic constraints. Here, we describe how linguistic constraint tools— represented for example by using a cultural artefact—can influence collaborative task performance (Sect. 3). We explain a new setting, where a marker is applied as a linguistic constraint tool based on a given cognitive representation model (Sect. 4). Based on our theoretical considerations we specify a research design based on a marker and complexity condition (Sect. 5). This design is embedded into a geographic map as the most intuitive cognitive representation model. Furthermore, we describe our collaborative task of identifying a target location, to evaluate the role of a shared marker in addition to a cognitive representation model. The described setting is a very common task, which enables participants to participate without any prior briefing necessary. The principle of least collaborative effort becomes continuously visible to the team by implementing delay discounting decision problem into the reward system. The setup becomes complete through the description of testing conditions in the manner of applied communication media and representation model constraints. Based on this specification, we describe the applied procedure in detail (Sect. 6). Central to our procedure is an implemented chat-tool integrated into a shared geographic map. While three participants are meant to solve the task at three working stations without any moderation, the tool provisions step by step the testing conditions and monitors the progress of a game round. The results show that we cannot observe the characteristics of team focused interaction (Sect. 7). With the first level of decision complexity, no real team interaction occurred. Based on a second level decision complexity, more intense team interaction occurred, but the marker condition achieves in general a disadvantage. Theoretically it is assumed that if participants collaborate they will be most successful if the discussion is constraint in some fashion (team focused interaction hypothesis) (Sect. 8). While our research

design tries to confirm the team focused interaction hypothesis, the results contradict
their assumptions. From our point of view, decision complexity seems an impor-
tant control parameter, which has not been covered with the given team focused
interaction hypothesis.

2 Contribution Model of Conversation

Tomasello (2014) states that one basic advantage humanity has is the capability and
motivation to collaborate and to help each other. Some human activities are only pos-
sible when multiple people are able to coordinate in a highly complex way (taking
for example "playing a duet on a piano"). The contribution model of conversation
contains a basic approach to explain how long the participants are interested in collab-
oration, or not. The model explains the coordinative behaviour of participants through
internal economic forces. This is based on the assumption that if people participate in
a conversation they act in a collaborative manner. Here, we summarise the basis of this
model, which is also explained in other contributions by Clark and Bangerter (2007),
Clark and Brennan (1991) and Clark and Schaefer (1989) (see also Fig. 1).
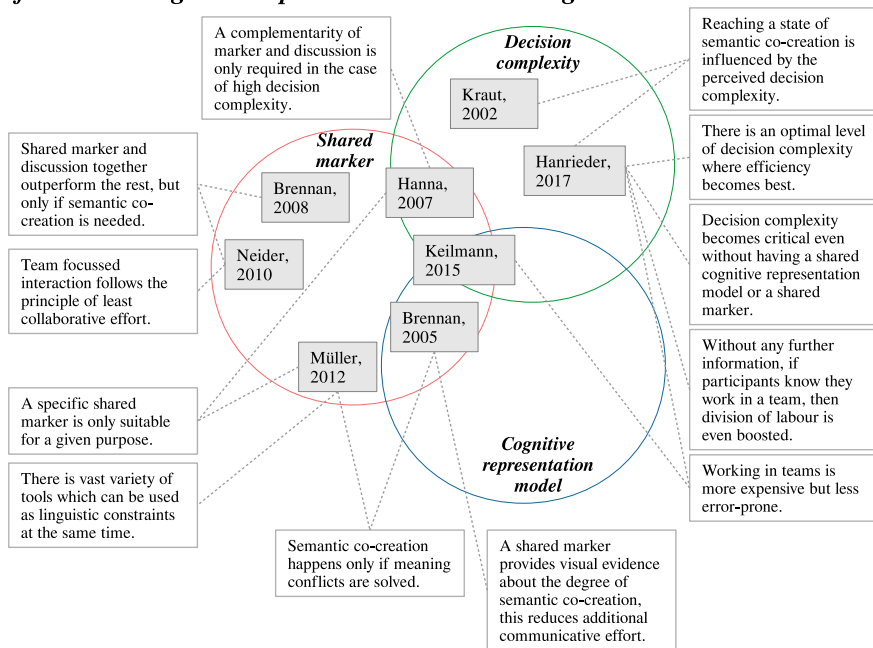

Fig. 1 Previous contributions: Team focused interaction for shared cognitive representation models
having a marker

Coordination, common ground, semantic co-creation, contribution: In collaboration through conversation people face the problem of coordination, which is implemented by using contributions in participatory acts (Clark and Schaefer 1989). In participatory acts, people act together, which requires them to synchronize in terms of content and timing. Taking musicians as an example, playing a duet on a piano. Both musicians have to confirm, which duet they would like to play (coordination of content) and while they are playing they have to synchronize their entrances and exits (temporal coordination). To enable people to coordinate in conversation efficiently, they need to build a form of common ground. Common ground can be understood as an invisible form of cognitive representation, which all participants accept. In communication, common ground cannot be properly updated without a process. The question is how to evolve an individual idea or conception of something into a form of community-wide accepted semantic co-creation, that is manifested by using a state of common ground (Gergen 2009). Plain and simple, semantic co-creation is given if all task-related participants have got a sufficient idea of how to solve a problem in a collaborative manner, successfully (Raczaszek-Leonardi and Kelso 2008). Participants achieve successful coordination when they reach two degrees of semantic co-creation: grounding and identification (Clark and Wilkes-Gibbs 1986). In identification, participant one tries to get another participant to pick out an entity by using a particular description. Identification happens as soon as the pick out behaviour of the second participant is visible to participant one. In contrast, grounding happens when both participants think that they have identified the correct entity. This means the entity has already been added to the participants' common ground. To put it in a nutshell, the required description to reach a common ground (content specification) and semantic co-creation form a unit of discourse, so called contribution (Clark and Schaefer 1989).

Reaching the grounding criterion using the least collaborative effort: In order to evaluate their conversation, the participants have to set a grounding criterion. A criterion that participant A was successful in describing, could be given if participant B takes the correct object. The grounding criterion is achieved, if a certain amount of effort was provided by the participants to reach a sufficient degree of confidence in the success of a communicative act with a specific purpose. In context of a given communication purpose, the grounding criterion is achieved when all participants believe that they have sufficiently understood (Clark and Schaefer 1989). The participants try to reach the grounding criterion with the least collaborative effort. They are motivated to minimize their amount of work by providing dialog contributions that are as efficient as possible. The concept of least effort was classically described by Grice's saying for quantity and manner (Grice et al. 1975). Grice's saying for quantity states: "Make your contribution as informative as is required; do not make your contribution more informative than is required". While for manner: "Be brief (avoid unnecessary prolixity)." If a contribution follows these sayings it is considered proper, that means the participants believe a contribution will be readily and fully understood by their addressees (Clark and Brennan 1991). Nevertheless, the principle of least effort does not make any exceptions for time pressure, errors or grounding (Clark and Wilkes-Gibbs 1986). For example, when under pressure participants may not be able to plan

well-formulated brief statements and in such cases the model of least effort fails. To overcome these problems the principle of least collaborative effort was formulated by Clark and Wilkes-Gibbs (1986) as follows: "In conversation, the participants try to minimize their collaborative effort—the work that both do from the initiation of each contribution to its mutual acceptance." In participatory acts the participants have to reach the grounding criterion, while minimizing their effort, this is characteristic of conversation in general.

Contributions as a historical process: Ongoing contributions of a discourse have to be considered in historical fashion (Clark et al. 2007). In a classical referential communication task by Krauss and Weinheimer (Krauss and Weinheimer 1964), a participant has to describe his partner which of the four presented abstract figures needs to be selected. To identify the correct figure, each team requires a number of descriptions and related feedbacks. The results of this experiments confirm that ongoing user interaction leads to coordination as a historical process; meanwhile the common ground is constantly emerging. Therefore, as interaction is continuing, descriptions become even shorter ((Krauss and Weinheimer 1964): e.g. (1) "the upside-down martini glass in a wire stand", (2) "the inverted martini glass", (3) "the martini glass" and (4) "the martini") and the number of required turns decreases over time (Clark and Wilkes-Gibbs 1986). In order to make descriptions gradually become more efficient, a form of functioning interaction requires some kind of working user interaction. The average length of descriptions only drops if participants can give direct feedback.

3   The Influence of Linguistic Constraint Tools on Reaching the Grounding Criterion

Bias on reaching the grounding criterion: The previous section demonstrated that reaching the grounding criterion is fundamental for communicative success. For this reason, it is important to understand how reaching the grounding criterion can be influenced. Answering this question is about looking for "tools", which are used to let semantic co-creation happen. Modifying the performance of this tools influences the success on reaching the grounding criterion. Three types of tools are required for reaching the grounding criterion: These are signs, practices and a communication channel. We follow the notion of Löbler (2010), while a sign is "everything which is perceivable, everything we become aware through the senses."; and further a practice "coordinate ways of doing and sayings." Further he noted that: "Practices are implicitly behind all forms of explicit coordination, they coordinate implicitly, and we can become aware of them by the ways we do or say things." If signs and practices wants to be applied, then a communication channel has to be used to overcome a spatial distance in the shared environment. Here we follow the basic channel notion of Shannon's sender-receiver-model (Shannon 1948): "The channel is merely the medium used to transmit the signal from transmitter to receiver."

It has been already pointed out, that practices, as well as the resources available within a communication medium, are termed critical factors (Clark and Brennan 1991). Reaching the grounding criterion can easily translated as what needs to be understood with a given purpose (Grice et al. 1975). This criterion changes through the application of a specific content practice suitable for a given purpose (Clark and Wilkes-Gibbs 1986). For example participants has to identify objects, then a conversation focus on them and their identities. The applied content practice has to ensure that the objects can be identified securely and quickly. Based indicative gestures as an exemplatory practice, an object identified if a speaker refers to an object nearby and the addressee can identify them by pointing, looking or touching. A communication medium (like e-mail or fax) got also an effect on reaching the grounding criterion, while their fulfilling of communication channel constraints differs (Clark and Wilkes-Gibbs 1986). There is a set of costs (e.g. formulation costs or understanding costs) that can quantify these constraints from different perspectives. Nevertheless the influence of signs has been not respected, yet.

Constraints: In this study we follow the idea of linguistic constraints by Pattee (1997) reformulated by Rascazek-Leonardi and Kelso (2008). A fundamental premise of Pattee's theory of living organisms states that there is an interelation between measurement and control. Here, control is about producing a desirable behaviour in a physical system by imposing additional forces or constraints. These constraints are not fixed, but are applied and adapted based on the demands of the environment. They are applied dynamically, following the purpose of a coordinated action.

Linguistic constraints: The described notion of constraints is limited to a specific moment and place. In addition to control, measurement is a symbolic result of the dynamic process. While constraints in a moment of control was limited to a certain point in time and space, the emerging linguistic constraints in the momement of measurement are not fixed. Linguistic constraints—instantiated through symbols—encode stable patterns of dynamic variables that are relevant to control something between an individual and some environment. The human's task of measurement is to choose a relevant pattern and ascribing a symbol to it. Together, linguistic constraints applied in measurement and control can only be understood in a given situation and context of a given space and time they are applied. They are covering the history of constraint application in language use based on multiple timescales.

Linguistic constraint tools: Bringing linguistic constraints into practice we have to notice that typically they are embedded. Hence, Löbler pointed out that "signs render services in helping to find what we are looking for (Löbler 2010)." For us it follows that linguistic constraint tools instantiate services in relation to linguistic constraints. These tools e.g. discussion, cultural artifacts, cognitive representation models or marker can be valuable to achieve a state of semantic co-creation more easily.

4 Using a Marker in Shared Cognitive Representation Models as a Linguistic Constraint Tool

In the last section we introduced the concept of linguistic constraint tool. The question remains open how linguistic constraint tools can help in achieving semantic co-creation based on shared cognitive representation models. In this section we present notion of team focused interaction and summarize previous findings on using a marker in cognitive representation models as an example of a linguistic constraint tool.

Origins of team focused interaction: Team focused interaction describes an approach to identify the correct target in situations of high decision complexity (e.g. identify a target from many). The hypothesis was introduced by Zubek et al. (2016). They evaluated the constraining role of cultural artefacts on the performance of a collaborative language task within a real world setting. The authors used a wine-identification task. Participants were separated into pairs and single probands. In contrast with single probands, pairs can talk freely to each other. Namely, they tried to identify wines, based on their shared tasting experience. Every pair has to talk about smell experience of wine, so it can be assumed that based on the same purpose, there might be similar practices applied. From external, the conditions of communication for pairs were the same. They can talk freely to each other, as long as they want in place they shared physically.

The cultural artefact was a wine tasting card that contains 21 items including a category and their available attributes in the field of taste, smell and general characteristics of wine. For example there was category "Alcohol" and an attribute "Light". In team interaction, the participants can use these taxonomy to describe their tasting experience. Their experiment was designed to evaluate the identification performance based on two conditions: the use of a wine tasting card and whether one participant uses this card or whether two participants use the card and interact freely.

The results showed that interacting pairs were better in identifying the correct wine than an individual wine taster. With the help of a wine tasting card the accuracy of individual participants did not improved significantly. The best performance was achieved, when a pair of wine tasters used a wine tasting card. Pairs using a card had more consistent vocabulary, than those without. The more consistent their vocabulary, the more they were successful in identifying the correct wine. In addition, participant pairs using a wine tasting card had a lower variance in their identification of wines compared to participant pairs without such cards. The lower variance relates to the usefulness of wine categories within the wine tasting card. These linguistic categories likely function as linguistic constraints by focusing the communication, making wine identification more reliable and precise. Together, a linguistic constraint tool can stimulate team interaction towards more focused communication, we name that idea the fundamental premise of team focused interaction.

Nevertheless, this study did not use a common cognitive representation model and even shared marker was not present within them. In both cases, the cognitive representation model was created by each participant, separately in their minds.

There are a couple of referring expression tasks, which cover our research interest as combination of having a shared marker, by using a cognitive representation model and with respect to decision complexity (see also Fig. 1).

Kraut et al. (2002): This study by Kraut and colleagues investigates the role of decision complexity on the communicative success. Decision complexity controlled by the puzzle difficulty (easy: non-overlapping, complex: overlapping elements) and the color drift (easy: static colors, complex: changing colors). These two measures are contrasts with respect to having a shared space, delayed (3 s delay) or not delayed. For evaluation purposes, a team of participants has to arrange a puzzle in the correct order. One participant explains to another how he has to re-order the elements. The results show that teams become faster in solving the task when the puzzle uses non-overlapping elements and static colors were present. Especially, changing colors become a problem to the participants if the screens are not shared immediately. Timing the utterances in discussion moment-by-moment becomes even more complex, because utterances meant to achieve semantic co-creation are biased. In this respect, it becomes obvious that achieving semantic co-creation is influenced based on the perceived decision complexity.

Brennan (2005): Brennan wants to observe the convergence of semantic co-creation moment-by-moment. She is especially interested in how the fact of having a shared marker as a visual indication can influence this process. The research design compares a factor of visual evidence (having a shared marker or not) in contrast with map familiarity (being familiar with a map or not). Based on a geographical map, a participant has to get to an unknown target location, which is visible and explained by a second participant. The participant's mouse movement gives evidence of what has been understood based on the speech (description) of the first participant. Brennan defines several time stages of comprehension towards semantic co-creation: start, first move, close to the target, reliably understood but not at the target, pause and identified. The results show that participants who have no shared marker not only require more time to finish a task, they also require a lot more words to get from a reliable to a final state, and they require a lot of time at the pause stage to get a collaborative acknowledgement. To sum up, a shared marker acts as a visual linguistic constraint tool that simplifies the process of semantic co-creation. Participants become more efficient when a shared marker is present. A second observation shows that solving decision conflicts becomes much more difficult without a shared marker. The final acceptance phase requires much more time than with the presence of a shared marker. Regarding this, decision conflicts can be present as a natural part of semantic co-creation.

Hanna and Brennan (2007): The authors ask if coordination by shared gaze can outperform speech. In the previous study, Brennan has designed shared gaze as something that gives visual evidence. In this investigation shared gaze is introduced as a new type of linguistic constraint tool in addition to discussion based on speech. The role of shared gaze in discussion is evaluated through structural task properties (the orientation of available elements and the distance of a competitor in relation to the target) that control the perceived decision complexity. The task requires to identify the correct target element from a set of forms that are presented in front

of the participant. A second participant knows the correct element and describe this target from another side of the table. Both participants are recorded by eye-tracking and voice recording. For evaluation purposes both recordings are integrated into one stream. The results show that eye gaze produced by a speaker can be used by an addressee to resolve a temporary ambiguity, and it can be used early. Shared gaze outperforms speech because its orientation becomes even faster. This observation let us conclude that there is a competition between shared gaze and discussion, which is won by shared gaze.

Nevertheless, the results hold true only up to a certain level of decision complexity. Shared gaze outperforms speech only in cases with no similar objects (no-competitor condition) or far distant similar objects (far competitor condition) in the possible answer set. In the case when there are similar objects close to each other (near competitor condition), no advantage of shared gaze in contrast with speech can be observed. That means a complementarity of multiple linguistic constraint tools as already described for multiple timescales is needed in cases of high decision complexity. In this respect the complementarity happens in time (multiple timescales) and space (multiple tools) as well.

Brennan et al. (2008): In a further study Brennan is interested in collaborative search scenarios in which both of the two participants are not aware about the target. The collaborative search was evaluated under a shared-gaze, shared-speech, shared-gaze and shared-speech or no sharing conditions. In their study participants had to identify a possibly present O within a set of Qs (O-in-Qs search task). Where participants searched together without sharing anything, accuracy was very low, in fact accuracy was lower than where individuals searched alone. Under shared-gaze conditions the best results, in terms of search duration and accuracy, were achieved. A longer search was observed in teams under conditions of shared-speech or shared-speech and shared-gaze. That observation that shared-gaze outperforms shared-gaze and shared-voice is obvious. The given task is clear to participants without any further negotiation to finish them successfully. If no semantics needs to emerge from collaboration, then no semantic co-creation is required. Even so, team focused interaction is not only about collaborating users having a linguistic constraint tool, it contains also a successful coordination based on semantic co-creation. That means having a shared marker (e.g. shared gaze) can outperform a combination of shared marker and discussion, but only in cases where semantics does not have to emerge.

Neider et al. (2010): Based on the previous study of Brennan et al. (2008), Neider et al. are interested in collaborative search scenarios with two participants working as novices. In contrast, a scenario is studied where consensus between the participants is required. Hence, the study requires that both participants together have to identify the correct target to finish them in time-critical manner. The study design compares shared gaze only, speech only, and shared gaze plus speech. In addition, a no communication condition is applied. In a sniper task, a virtual environment is used to identify the correct sniper target together. The results show that shared gaze together with discussion outperforms shared gaze alone. This observation confirms the previous assumption that semantic co-creation needs to be required to unfold the benefit of team focused interaction.

Further, Neider et al. evaluated that in these cases the principle of least collaborative effort becomes true. Having shared-gaze in contrast with speech only condition, the first participant was faster in identifying the target location because its location doesn't need to be described in detail. If the situation becomes clear to the second participant the first participant only has to note that the second has to go to the target and the task was solved successfully. In a not so clear situation, monitoring gaze behavior as well as more scenic descriptions slow down the consensus phase. Two scenarios with different costs on consensus side are observed. Only when necessary the participants dive into a more detailed discussion. Based on the principle of least collaborative effort such a behaviour can be expected.

Müller et al. (2013): Investigates the role of discussion context and the question if shared mouse can approximate shared-gaze behaviour. They applied a puzzle arrangement task and distinguished between sharing a common gaze, a common gaze and speech, sharing a mouse and speech or speech only. A participant had to arrange the correct order of a puzzle from a set of randomly organized puzzle elements based on the description of a second participant. In addition to the different sharing conditions, one group of participants had to strictly follow particular instructions (low autonomy), while another group could rearrange the puzzle freely (high autonomy). The level of autonomy showed the strongest effect for all communication conditions. Low autonomy conditions resulted in better task-performance, based on lower error rates, independent of the communication conditions. This shows that in a given task-context, more specifically having high or low autonomy, acts as a constraint in communication. The results of Mueller and colleagues underlines the observation that interacting-pairs perform best when they are restricted by some form of linguistic constraint tool. In contrast to Zubek et al. (2016), Muellers study did not require a predefined taxonomy—like a winetasting card—in order to enforce a specific behaviour of the participants. Furthermore, having understood autonomy as a discussion rule tool is no shared marker. Benefits and costs are quite different. Based on such a discussion rule, e.g. continuous monitoring as described by Neider (2010) is not needed. The results also show that in cases of low autonomy, the shared gaze and discussion condition perform even better than discussion condition alone. That means several linguistic constraint tools can be used at the same time. Discussion, shared marker, shared taxonomy or even conversation rules are only some examples of such tools.

By comparing shared mouse and shared gaze, it was additionally observed that shared mouse becomes a good approximation of shared gaze to the given task. Solution times were within the same range and error rates were only higher for gaze than for mouse transfer when the former was used without a speech channel. Considering shared mouse a visual indication requires much more time then shared gaze, but in contrast shared gaze provides much more marker data which has to be interpreted by the participant in a sufficient way. Summarizing, linguistic constraint tools are more or less suitable based on the current purpose of coordination.

Keilmann et al. (2017)[1]: In our terms, Keilmann and colleagues try to evaluate the team focused interaction hypothesis, where a cognitive representation model becomes a linguistic constraint tool. They compared a partly visible and a completely visible labyrinth to one another, either with an individual or with participant pairs. The study examined having a labyrinth as an example of a cognitive representation model or not and whether participants work in a team or alone. If the labyrinth was shared, both participants could see the complete map. In contrast, if it was not shared only a specific subarea of the labyrinth was visible to each participant, individually. The current position of a participant as shared marker was only shared if a participant was located in the visual field of the other. As a third factor, the perceived decision complexity was controlled based on the number of intersections present in the labyrinth. Communication between the participants in collaboration was allowed via headphones. As fast as possible, the participants have to search the complete labyrinth to get all pickup items. The results show that collaborating participants in contrast to an individual searcher are faster and require less trajectory lengths, even though they get higher error-rates in picking up correct items. Hence, team interaction is more expensive than searching alone, but together pairs can achieve a better identification performance. These observations refine the idea of Zubek et al. (2016) that team interaction improves the identification accuracy but requires more communication costs to achieve a coordinated behavior. If the cognitive representation model was shared teams generally outperform single participants. We consider the labyrinth as an example of a cognitive representation model, which is another linguistic constraint tool. Using a shared cognitive representation model collaboratively enables interactions to be more focused.

Hanriede (2017): Hanrieder transforms the Keilmann's stimuli from a top-down into a within-environment view. The participants in the role of firefighters search a floor for casualties as fast as possible. The given task was applied with two cooperation modes (either as individuals or pairs) and several levels of labyrinth complexity (8, 11, 14, 17, or 20 intersections per environment). Hanrieder's setup comes very close to Keilmann's, but in contrast it provides no shared cognitive representation model and it prohibits communication between the participants, who work in teams. Beside the mode of cooperation (individual vs. team collaboration), the decision complexity is controlled based on the number of intersections. It can be shown that the number of intersections have got a negative impact on the task performance. The participants be it individuals or pairs, needed less time to finish, travelled shorter distances and got less error-rates (missed less pickup items). More detailed than Kraut, 2002, it was possible to observe that an increasing decision complexity leads to higher costs and error rates. Keilmann's observation that groups in contrast with individuals are more expensive, while the error-rates are on a lower level which can be confirmed also in a virtual environment.

---

[1]Note: The contribution by Keilmann et al. does not seem to be available for public use. Hence, our description is based on explanations by Hanrieder (2017).

Additionally, the degree in division of labour is measured by the self-overlap of the participants (number of locations at the labyrinth which have been visited more than once) in comparison between individual and pairs. The results show that individual and pairs achieve less overlap for more complex environments. Comparing pairs to individuals, pairs achieve much less self-overlap than individuals. With regard to team focused interaction, this insight is quite interesting because division of labour can be improved even if the participants cannot communicate and have no shared tool working as a linguistic constraint. This observation does not contradict the fundamental premise of team focused interaction. Division of labour seems to be a fundamental practice which is enhanced by team focused interaction.

Prediction: In the following, we want to predict the effect of a shared marker if it is used on top of a cognitive representation model. Such a marker is a very flexible user-driven linguistic constraint tool, which is embedded into a shared cognitive representation model. A marker used in addition to a shared cognitive representation model limits the decision space. By using a marker each participant can time their words and actions better. At any moment of content specification and semantic co-creation, the participants exchange evidence via the shared cognitive representation model, whether the grounding criterion is fulfilled or not. In addition, a marker position informs all participants how far they are from reaching the grounding criterion. If the marker was not moved, then only the given cognitive representation model can be used to limit the decision space.

Using a shared marker is a very common setting in the presented referring expression studies. If such a shared marker is present then each case using them is enforced. The participants have to move the shared marker to fullfil the task successfully. Such an "enforced move requirement" is a problem because it prevents a fair competition between the shared marker and discussion as two ongoing linguistic contraint tools. We are interested in the question if the fundamental premise of team focused interaction becomes true, even if shared marker usage becomes optionally. If we compare experiment durations of previous studies, than we can observe that study durations can be grouped in two categories. The first category experiments are those having a total duration up to 20 s, while the second category observes durations up to 140 s. If we are think about the nature of discussion, then we think that coordination in first category of tasks is very straight-forward. Neider et al. (2017) named that phenomenon one feedback based on description. From our point that means that there is a communication channel but no real team interaction occurs. Such a behaviour can be explained because of a very low perceived decision complexity. Hence we predict that if decision complexity becomes very low than no real team interaction occurs. With second higher level of perceived decision complexity the fundamental premise team focused interaction should become true. Having shared marker should provide an advantage to achieve a good identification performance (Table 1).

Table 1 Comparing previous studies based on the observed task duration

Total durations (approx. mean values)

Contribution	MIN	MAX
Kraut et al. (2002)	40s	120s
Brennan (2005)	10s	20s
Hanna and Brennan (2007)	n.a.	n.a.
Brennan et al. (2008)	30s	120s
Neider et al. (2010)	8s	12s
Müller et al. (2013)	50s	140s
Keilmann et al. (2017)	n.a.	n.a.
Hanrieder (2017)	50s	120s

5 Setup

The team focused interaction hypothesis claims that using a linguistic constraint tool influences the success of reaching the grounding criterion. In our study we are specifically interested in using a cognitive representation model as a shared artefact, with an additional but optional shared marker available. In collaborative manner the team can solve the given task only based on linguistic features through discussion. It is up to the team to use a shared marker as visual support. The limiting power of a marker will be evaluated by comparing groups with and without a shared marker. The complexity of the cognitive representation model appears to strongly impact the performance of such markers. Hence, we implemented marker conditions with two degrees of complexity.

Which cognitive representation model is suitable for the given evaluation purposes? In our study, a cognitive representation model is present as a shared artefact. There are several forms of cognitive representation models (such as the conceptual space (Gärdenfors 2004), the biplot (Gower et al. 2011) or the associative semantic network (Collins and Loftus 1975)), which differ in their representation (e.g. spatial vs. graph representation) and in their dimensionality and the number of entities they consist of. However, we only applied the geographic map (Monmonier 2018) as a very intuitive example of a cognitive representation model. For our study we required participants to understand the model without any prior learning effort, as such we selected the geographical map as the most widely accepted model. Geographical maps represent complex structures based on standardized criteria (typically distances). A map can be specified as 2-dimensional ($l \times l$ e.g. a map of Germany) or 3-dimensional length space $l \times l \times l$ e.g. an orbit map of our galaxies) By using a globally standardized metric to describe the orientation between very many entities within a space (e.g. all cities in a country), it becomes possible that a large society of people can coordinate within this shared space. For example, millions of deliveries are shipped about the whole world every day only based on one standardized geographic world map. These characteristics make a geographical map very useful where

a large group of people want to coordinate in a shared space (Monmonier 2018). We ensure our model preference by asking the user about the tool familiarity of some other promising cognitive representation models. We measure tool familiarity as the degree model usage within the daily life.

What form should the referring expression task take? We want setup a remote referring expression task in a shared environment, while such a task allows to observe if-based on a given referring expression an intended referent can be picked out (Clark et al. 2007). Such shared environment tasks are possible in two settings: First, the expert-novice setting (e.g. Müller et al. 2013; Brennan 2005), here one participant is familiar with the target (expert), while another participant who is not familiar with it (novice), has to identify it. Second, the novice-novice setting (e.g. Brennan et al. 2008; Zubek et al. 2016) is about identifying a target, while the target is unknown to both of the participants (both act as novices). To evaluate the impact of shared markers on reaching semantic co-creation, it is required that shared markers can play a primary role for coordination purposes between the participants. Hence, we prefer to setup an expert-novice setting, because in such a setting, participants have to collaborate to identify the target. Nevertheless, in novice-novice settings it is possible to search separately (Müller et al. 2013).

The aim of our evaluation is to observe when and how semantic co-creation occurs within a group of participants. We decided to set-up a group of three participants, a describer, an actor and an observer. Under shared marker conditions the marker is visible to all participants, whilst under non-shared conditions the marker is not visible to the participant whose task it is to describe. In such conditions it is possible that the actor (participant carrying out actions) can help the observer (passive participant observing interactions between the other participants) by using the marker.

The task itself should be implementable based on a geographic map as an example of a cognitive representation model. Here, the map task is one example, where one participant needs to explain the route of a map to a second participant (Anderson et al. 1991). The two participants are presented with the same map, the first participant is shown a route marked on the map and is asked to describe this route. The second participant marks down his comprehension of the route, based on the given description. The map task differs from other tasks as the communicative success is measured on a metric scale. Describing a route within a map is a complex task, which requires high intellectual effort of the pair involved. In our study we implement an easier target location task, similar to that used by Brennan (2005). In Brennan's target location task a car icon has to be manoeuvred towards a target location. Only the participant whose task it is to describe can see the target location for the car on their map. The actor tries to find the unknown target location by applying the instructions described to him. The actor can use the shared mouse to relocate the car icon within the geographic map based on the hints given. Sharing mouse-movement is evaluated to uncover the current state of comprehension towards the grounding criterion, continuously. In the study by Brennan (2005) the task is completed once the actor places the car icon very close to the target. Such a setting forces the actor to use the car as an existing marker. However, we want to make the use of the markers optional in

order to evaluate the benefit of using them. Hence, our task ends when one city is selected from a given list of all cities present on the map. In principal, it is possible to finish this task without moving the marker.

How are model considerations implemented within the task? The aim of our task is to make the characteristics of the contribution model visible to all participants at any given moment of interaction. As described previously, the contribution model is implicitly present while participants are interacting in conversation. Nevertheless, there are no defined conditions concerning how brief or detailed each contribution to conversation needs to be in order to be understood. This lack of specificity leads to a bias of incorrect reward assessment by the participants. Applying a time constraint to the collaborative task incorporates time pressure and makes participant contributions briefer (Neider et al. 2010). One disadvantage of such time constraints is that it is harder to interpret how much effort the participants invest in conversation. Hence, we prefer to describe the task the participants have to complete as a collaborative conflict situation of least collaborative effort to reach semantic co-creation.

The "conflict" becomes visible to the participants online through scores assigned to participants' actions based on the delay discounting decision problem (Scherbaum et al. 2016). In delay discounting decision problems, a single participant has got two options of which they have to select the most beneficial one. The first option is named sooner smaller (SS) option, which means the user can get this one very fast but he needs to accept a lower reward value. In contrast, the second option is named later larger (LL) option. This option returns a much greater value to the user, but it is much more difficult to reach it resulting in a long delay. Unlike in the classic single participant approach, multiple participants who are trying to coordinate try to reach the highest degree of value discounting. Based on an initial reward score value each team member has to ensure that their actions reduce the team score as little as possible, while reaching the grounding criterion should be achieved as quickly as possible. In our case, the describer needs to decide whether they want to apply a more detailed description (SS-option) or only slight hints about the location, e.g. using the words "hot" or "cold" (LL-option). If a participant applies his own description, he wants to ensure that he reaches the grounding criterion fast, even though only a small team discount can be achieved. In contrast, if a describer applies only hints, he tries to achieve a larger team discount, while it becomes more difficult for the team to reach the grounding criterion, because such hints are much less informative. In our case the actor and observer need to decide whether to select just a target subset (SS-option) or whether they want to know the exact location (LL-option). Actors or observers who only select a target subset slow down the required grounding criterion, because the correct answer needs only one element within the given subset. This option has the disadvantage that the team score decreases very much. In contrast, if an actor or observer selects a unique correct answer the requirement to reach the grounding criterion is much higher, because only one correct answer needs to be identified. This more delayed option seems charming because it decreases the team score much less. In applying participant action scores in the coordination task, it will be possible to measure the degree to which the SS and LL principles have been

used, interactively. For implementing collaborative delay discounting problems, we decided to implement the text chat tool instead of audio channel communication (e.g. Neider et al. 2010).

What are the test conditions? Test conditions of referring expression tasks can be described using the limitations of communication media (Clark and Brennan 1991) along with the content provided with descriptions or identification skills in order to determine the credibility of other participants (Edwards and Myers 2007). Co-occurrence assumes that the participants are present at the same time. Our task is applied to three participants, who have access to the same shared workspace in different roles. Together with a shared cognitive representation model the participants can communicate by using a shared chat system. Here, they can write and read messages at the same time (simultaneity), they can look at older messages within the chat protocol (reviewability) and read their messages before they submit them into the chat (rereading). In shared space scenarios communication delay becomes an additional critical issue (Kraut et al. 2002). The communication between the nodes based on LAN connection as well as our script performance happens without any perceivable delay. Furthermore, we need to ensure that the task is achieved based only on the geographical map and chat messages available to the participants. No additional communication media should be used (e.g. other messenger services), no other sources of information should be available (e.g. Wikipedia) and no common ground should exist between the participants before starting the task. To guarantee these test conditions, participants worked at prepared working stations, which only offered access to the testing environment. The use of mobile phones was prohibited during the evaluation.

Within the test conditions for the cognitive representation model it is considered the following questions: what do the participants already know about the map (pre-existing background knowledge (Brennan 2005)), how are the cities of the map structured (entity structure) and can the participants use a symbol for a city for communication purposes (symbol entity referencing). Pre-existing background knowledge occurs where participants have some common ground beyond the task, which could help them to complete the task more easily. If our task were to use a map of Germany and the participants were German they could use their background knowledge to identify particular places on a common map faster than if their task involved a map of an area unknown to them all, e.g. Ukraine. Our study, in fact, deploys maps of Ukraine and other countries for which we consider it unlikely that participants will have previous knowledge of. The second factor, entity structure, is about the complexity of coordination within the cognitive representation model. This complexity is indicated by the number of elements and the proportion of reference points from all elements (non-reference points). A reference point is a location with discriminable features, which allow a subject to have a geographical orientation (Sadalla et al. 1980). Having reference points improves orientation in cognitive representation models (Hanrieder 2017). Within most maps there are a set of reference points, in our case popular cities in a country (e.g. Kiev in Ukraine). If a reference point was a target the participants could refer directly to cities based only on their name (e.g. a city which is marked "Kiev"). To identify the correct target the describer could send

this description to the actor, who can identify this place easily, without any further interaction. To avoid such a behaviour, we add a set of randomly chosen less well-known cities, one of which needs to be identified. We introduce these random cities with a symbol instead of a name. This approach prevents town names from being used as descriptions. However, the describer could refer to a town's symbol instead ("go to the town x_1"). To solve this potential problem, we inform all participants that their symbols for the given towns are all different, making such references no longer useful. This should prevent the participants from using the symbols of towns for coordination purposes. It also simplifies the map, because only reference points can be used between the describer and the actor. An increasing number of reference points makes orientation on the map landscape easier. Just as well, identifying a target location become easier if the number of potential decision points is smaller. Low map complexity means there is a large number of reference points and a small number of non-reference points. We defined complexity level 1 (low) as consisting of 5 reference points and 10 non-reference points. Map complexity level 2 (high) comprising 1 reference point and 25 non-reference points.

6 Methods

As each task is limited to a duration of five minutes, each team completes the whole experiment (six trials) within 30 minutes. Both the actor and the observer should identify the target location which means there are two task results per task. In total twelve task results are recorded for each team. Resulting in a total of 156 instances for evaluation.

6.1 Participants

Our task was completed by 13 groups each consisting of three participants, with 6 trial rounds. In total there were 39 participants, of which 17 were female, the average age of participants was 32. All participants had normal vision, or their vision was corrected to normal with glasses or contact lenses. Each of the participants gave informed consent to take part in the study and received a natural gift (a bottle of water or a piece of fruit) after completing the experiment. Each team member of the winning team was given a bouquet of flowers as a gift.

6.2 Apparatus and Stimuli

Stimuli were presented on three laptops simultaneously, each with a normal RGB background on a 14.1-inch screen at a resolution of 1377×768 pixels with 60 Hz

refresh rate. In our evaluation, we use an individually implemented analytics pipeline using a survey tool and a pre-processing tool. The survey tool handled our specified task, we described in the last section. It was implemented using PHP and AngularJS and had an integrated MySQL relational database. The chat environment was based on Socket.io. The presented geographical map was implemented by D3 and TopoJSON in a similar way to the tutorial by Mike Bostock. Additionally, we implemented the Natural Earth dataset of GDAL to create each of the maps which included country polygons and populated cities within each country. The pre-processing tool sets up a database for survey data and transformed this data into a dataframe, which could then be directly evaluated using statistical analysis tools such as IBM SPSS Statistics.

The task user interface: The interface of the team workspace consists of a shared geographical map, a chat system, information on the current reward and remaining time and an area to apply participant interactions, with options such as "add a description" or "select target location" (see also Fig. 3). The describer, whose task it is to describe the target location, views the same geographical map as the other participants, but additionally one of the random cities highlighted. The describer can use two forms of participant interaction. They can describe the target location freely or use pre-defined hint buttons, such as short messages indicating "cold" (far away) or "hot" (close) within the chat system. The maximal message length for communication is 67 characters, comparable to the length of an SMS. Both interactions of the describer are reward related. A short hint ("cold" or "hot") relates the SS-option and
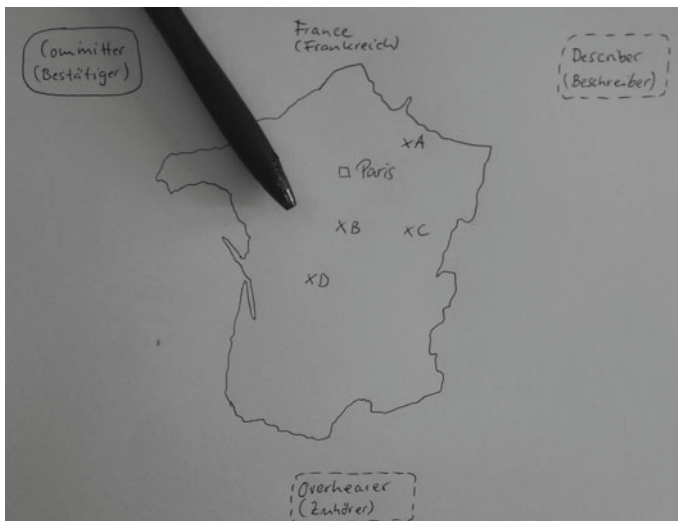

Fig. 2 *Paper prototype of our map task*: The paper prototype of the map task containing a simplified representation of France including Paris as a reference point (popular city) and four non-reference points (random cities), which are referred to via symbols. In preparation of the three team members, each participant of a team was positioned randomly around the map according to the roles described at the edges of the paper
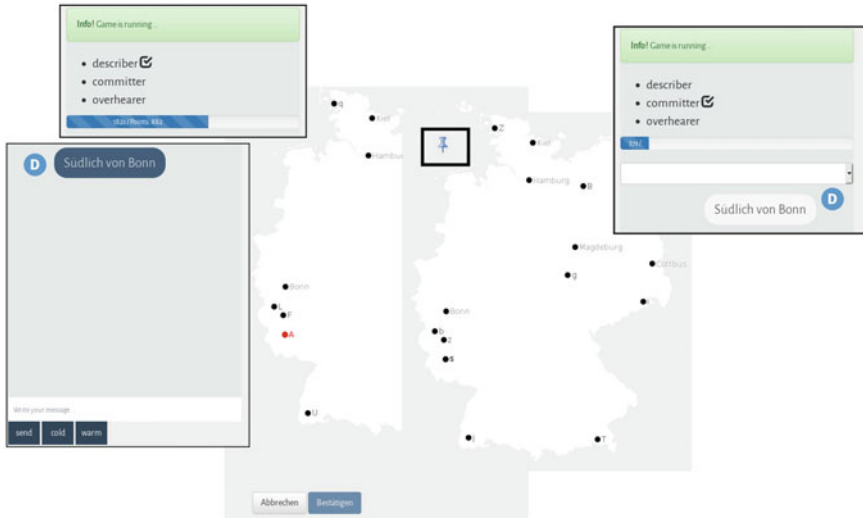
Fig. 3 *Target location task interface*: The user interface has a shared geographical map of a country like Germany containing a set of well-known cities (e.g. Bonn) and hidden cities (marked with symbols e.g. *b*). In the user interface windows on the left, those of the describer, town "A" is marked red, this is the city they have to describe the location of. The user interface windows on the right show that the actor sees the same cities but marked with different symbols. The describer needs to refer to the city of Bonn and indicate with a message that the target location is "south of Bonn". The actor can respond to this message by moving the marker (shown here as a pin above the map) or by replying to the message or selecting a reply option for a predefined answer set

reduces the reward by 10 points, sending a longer text message is the LL-option, which reduces the reward by 50 points.

The concept of collaboration: All participants are directly updated about communication methods used as they can all see the remaining reward amount. The actor can read the describer's messages and thus move the marker towards the potential target location. This marker is visible to all participants but can only be moved by the actor. Via the marker the actor can indicate where they assume the target location to be based on the describer's messages. The actor can also comment on any given explanation of the describer freely, without any costs regarding the reward. The actor has two options to complete the task: (a) by selecting the target location they assume to be correct (LL-option—select 1 city of 10 options in complexity level 1 or 25 options in complexity level 2) or (b) by selecting a subset of target locations, one of which should be the target location (SS-option—i.e. select 1 of 5 options, while each option represents in complexity level 1 two cities or in complexity level 2 five cities). Both of these options are also reward related: the LL-option results in a further reduction of the reward by 50 points, whereas the SS-option results in a 10 point reduction of the reward. In SS-options, cities are clustered based on their proximity by using same-size k-means clustering.

6.3 Procedure

Preparation: Before starting the task, a moderator explained the task and the participants of a team completed an initial test-round together by using a simplified paper prototype (see also Fig. 2). Besides, the participants had to evaluate how familiar they were with popular cognitive representation models. The moderator also informed the participants about the basic notion of these cognitive representation models by means of an example. The three participants carried out the task in separate rooms, each provided with a laptop. Communication among participants was only allowed within the provided chat room. Participants had to deposit their smartphones outside the room and the provided laptops had no internet access. After an introduction to the task by a moderator, each participant was left alone in their respective room with the laptop and the task without any further discussion with the other team members. Each participant had to sign-in to a shared workspace where they were then randomly assigned a role (describer, actor or observer).

The location identification task: With the shared workspace each participant within a team sees the same map of a country. This country map contains a set of reference points (popular cities of a country, e.g. Berlin, München, Hamburg, Frankfurt or Stuttgart for Germany) and a set of non-reference points. Non-reference points are cities which are labelled by a personally chosen random symbol (like a_1 or a_2) The team of participants begins the task together with a time limit of 5 minutes. The aim of the task is for the actor and the observer separately to identify the correct city as efficiently as possible, based on the hints given by the describer. The reward for achieving this task is 1000 points at the start of the game; this reward decreases as time passes and with increasing participant interactions. The team with the highest score after successfully completing the task wins. Unsuccessful teams who do not manage to complete the task end the game with 0 points. Additionally, the task ends earlier if the reward is reduced to 0 points based on the team's amount of participant interactions. After starting the task, the remaining time is displayed in the shared space, along with the current reward, which decreases at a rate of 1 point per second.

Finishing the task: Once the actor has correctly identified the target location by selecting the correct target location, all participants are informed via the chat that the actor has finished, the location is however not visible to the other participants. If the actor has completed the task first, they should help the observer to also find the correct target location. Hence, an actor who has finished stays on in a running game session and can write messages to the describer and move the flag. Giving answers is not possible any longer. Observers are by nature not allowed to interact with the shared space, but they can see everything which is happening. They can use additional user interactions of actor and describer to identify the correct location. Unlike the actor, the observer cannot interact and send chat messages or move the marker. The observer is truly just an observer who can see what the other participants are doing. In addition, she can apply an answer to finish her task. When the observer and the actor select the same target location they are rewarded with the same number of points. If the observer completes the task by correctly identifying the target location,

afterwards they are not able to give hints and can only continue to observe the other participants' behaviour. When both the actor and observer have completed the task, the team is rewarded with the current reward visible to them on their screens.

The marker condition: When only one participant has been able to complete the task the reward amount continues to decrease until the maximal task duration has been reached. The task is carried out under two conditions: either the marker is visible for the describer or not. Under "no marker" condition, the marker is still visible for the actor and the observer (and can be moved by the actor). Therefore, the marker is only helpful for the participants in these two roles.

6.4 Design

Each participant of a team is randomly assigned a defined role (describer *D*, actor *A* or observer *0*). In each role the task is applied either with the use of a marker (*M*) or without a marker *NM*. This combination of three different roles and two different marker conditions results in a total of 6 trials per team. For example, the following trial-order could be applied: (1) $O - NM$; (2) $D - M$; (3) $A - NM$; (4) $A - M$; (5) $O - M$; (6) $D - NM$. Each gameplay consisting of 6 rounds is based on geographical maps of the same complexity level. Whether a gameplay is based on complexity level one or two is assigned randomly.

A new geographical map was generated for each trial, from a set of the 6 countries. Each country contained more than 100 possible cities.[2] For each country only ten cities were selected as candidate reference points (popular cities), the rest were categorised as potential non-reference points (random cities), which were also randomly selected.

7 Results

Our initial focus was the use of the cognitive representation model. To confirm the suitability of the geographical map as a preferential cognitive representation model we asked participants to assess their usage of four cognitive representation model options (a geographical map, a biplot, conceptual space and a semantic network) in their daily life. The results, based on a 7-point likert-scale (from (1) "I don't know what this is" to (7) "I am using it regularly in my daily life") are shown in Fig. 4. The results reveal that the conceptual space and the biplot were the most unknown representation models. 23 of 39 participants didn't know what conceptual space was or had never used it. Similarly, biplots had never been used by 24 of 39

---

[2]Mexico incl. 1190 cities, Norway incl. 417 cities, Philippines incl. 156 cities, Puerto Rico incl. 242 cities, Portugal incl. 286 cities, Ukraine incl. 510 cities.
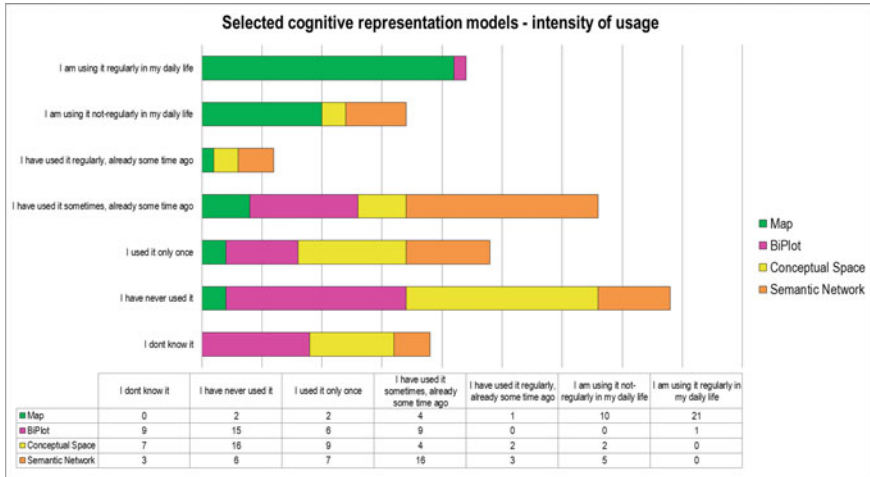
Fig. 4 *Intensity of usage*: 39 Participants evaluate how intensively they use four types of cognitive representation model: a geographical map, biplot, conceptual space and semantic network. The diagram shows how often the selected options (ranging from "I don't know what this is" to "I am using it regularly in my daily life.") were chosen for each cognitive representation model

participants. In contrast, semantic networks were more well-known and used by a larger proportion of participants. Of 39 participants, 24 stated that they used semantic networks, here responses ranged from "I have used it sometimes, but some time ago" to "I use it, but not regularly in my daily life". Nevertheless, the geographical map was evaluated as the most widely used cognitive representation model. 31 of 39 participants confirmed that they used geographical maps, even though not regularly in their daily lives. Based on these results we consider the geographical map as a suitable cognitive representation model for our study purposes, as it can easily be used by a broad range of participants.

We also wanted to evaluate whether the complexity of a geographical map influences the level of interactivity used to complete the task. Based on the principle of least collaborative effort, interaction itself contains the application of linguistic constraint tools which are used more intensely in complex situations. We hypothesised that if the complexity of a cognitive representation model is too low, then no team interaction emerges. To investigate this, we compared the two levels of map complexity. Complexity should influence all levels of interactivity, which describe the nature of a task round. As such we evaluated several indicators of interactivity: how often answers were given, the number of messages the describer sent and how often the actor responded to a message or moved the marker. Table 2 lists these indicators of interactivity. Using the complex geographical map, the describer had to send considerably more long messages (53.3%) and short messages (60.0%) than under simpler map complexity conditions. Where initial descriptions could not narrow down the target location enough, multiple long messages were required. Short messages used

Table 2 The relationship between complexity and interactivity: Several indicators of team inter-action are compared with two levels of complexity. Complexity level 1 (2) contains 5 (1) reference points and 10 (25) non-reference points. Complexity level 2 compared with level 1 requires a much higher degree of interactivity

	Complexity level...		
Interactivity indicator	1 (5–10)	2 (1–25)	p-value
Describer			
Sends two or more long messages	29.6%	53.3%	0.006
Gives short feedback (e.g. "hot" or "cold")	12.5%	60.0%	0.000
Actor			
Moves the marker at least three times	11.1%	37.8%	0.000
Sends two or more comments	11.1%	44.4%	0.000
Actor and observer			
Answers more than once (single-answer-option)	0.0%	24.4%	0.000
Answers using the answer set	0.0%	34.4%	0.000

by describers under these conditions tended to be small hints relating to previous interactions with the actor. Under complex map conditions (complexity level 2) the actor used the option of moving the marker more often (37.8% more than under less complex map conditions) and responded to the describer more frequently. An average of 44.4% of all actors gave feedback with at least two comments. Under map complexity level 1 participants selected the correct target city, rather than a subset of target cities. Comparing the map complexity levels there was a significant difference between the number of long messages sent by describers ($p < 0.01$), responses by actors ($p < 0.01$), movements of the marker by actors ($p < 0.01$) under the two map complexity conditions. Overall, there is a significant difference in the degree of team interaction required to complete the task between complexity level 1 and level 2. Figure 5 visualizes these differences based on two session examples. The results show that in contrast to complexity level 2, complexity level 1 requires very little team interaction to solve the task. Only when complexity increases does the level of interaction become more intense.

We also assessed the influence of the marker and the degree of interactivity on communicative success. From our previous observation we conclude that complexity (as a control parameter) influences the nature of interactivity. Hence, we focused on trials with map complexity level 2, where the participants appeared to be under higher pressure to interact. We used our results to evaluate the interactivity hypothesis, initially observed by Zubek (2016): When a pair of participants try to identify a target, they perform better than where one participant attempts this task alone. We reformulate this statement for our study: When a pair of participants interact intensively, their performance (in terms of task completion) is better than when interaction is very limited.
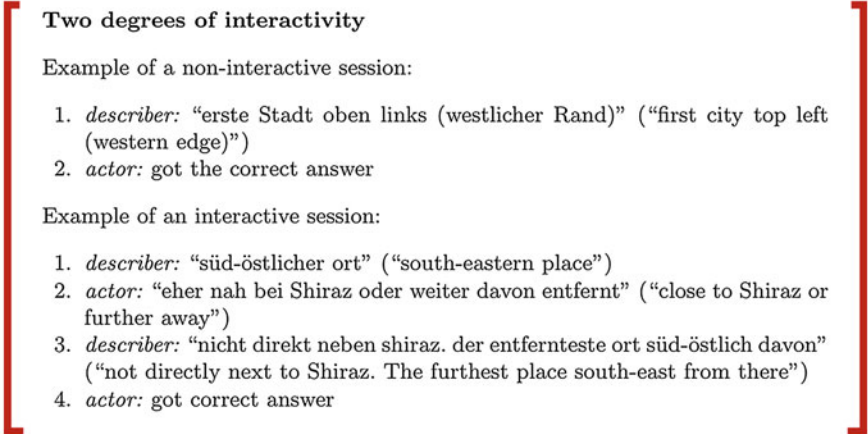
Two degrees of interactivity

Example of a non-interactive session:

1. *describer:* "erste Stadt oben links (westlicher Rand)" ("first city top left (western edge)")
2. *actor:* got the correct answer

Example of an interactive session:

1. *describer:* "süd-östlicher ort" ("south-eastern place")
2. *actor:* "eher nah bei Shiraz oder weiter davon entfernt" ("close to Shiraz or further away")
3. *describer:* "nicht direkt neben shiraz. der entfernteste ort süd-östlich davon" ("not directly next to Shiraz. The furthest place south-east from there")
4. *actor:* got correct answer

Fig. 5 *Two degrees of interaction*: In the first example, the describer sends one message and the actor was able to identify the correct target based solely on this description. Contrastingly in the second example further interaction and refinements are required to select the correct target. It becomes apparent that interaction only emerges when an initial linguistic constraint tool—the message sent by the describer—is not sufficient to reach semantic co-creation (to complete the task)

While we focused on cases where map complexity level 2 was used, interactivity was measured containing only most distributed variables. Most distributed variables are those having the biggest diversity of observed values. In our results, the number of comments made by the actor and the number of movements of the marker were the measures of interaction with the greatest variability. We compared these measurements with indicators of communicative success. The main indicator of communicative success is how many participants successfully completed the task. Here, the answer can be two (the actor and observer), one (either the actor or the observer) or none (neither the actor nor the observer). Further indicators of communicative success are the time taken to complete the task, for the first and second participant in each team to finish. The results for these indicators are summarized in Table 3. Based on these results the hypothesis that participant pairs perform better when interacting more must be rejected. Of all teams requiring 0 comments for finishing a task, 86.5% were successful. Likewise, of all teams that required 2 comments or more, only 40 % were successful. This observation is underlined when we look at the duration when the first participant (be it actor or observer) was able to finish the task. Teams which required no (vs. two or more) comment for finishing a task, were able in 91.7% (vs. 32.5%) to finish a task in fastest half of the first completion times, successfully. Same is true when we look at the number of marker movements. Teams which moved the flag 0 times (vs. 3 or more times) were in 85.0% (vs. 29.4%) under the fastest half of first completion times. Summarizing we can state that the best performing part of a team performs better if they do not interact intensely. Best performing part in this sense means two of three participants having one identifying participant, who was able to complete the task faster. Here, no benefit of interaction on communicative

Table 3 *The relationship between participant interactions and communicative success*: Two indicators of communicative success are compared with two measures of the degree of interaction. It can be summarized that the participants were most successful if they did not interact. However, it has become obvious that interaction is helpful in a way that all participants finish the task successfully

Indicator of communicative success	Number of comments used			Number of marker movements		
Degree of...	0	≥ 2	p-value	0	≥ 3	p-value
All participants successfully completing the task	86.5%	40.0%	0.004	75.0%	44.1%	0.119
Fastest 50% of teams completing the task	12.5%	60.0%	0.000	15.0%	61.8%	0.001
Fastest 50% of participants completing the task	91.7%	32.5%	0.000	85.0%	29.4%	0.000

success can be observed. Nevertheless, the results also make clear that the final completion was improved by a high degree of interaction. Teams which required no (vs. two or more) comment for finishing a task were only at a level of 12.5% (vs. 60.0%) under the fastest half of participants to the full completion of the task. Similarly, of teams which did not (vs. 3 or more) move the marker, 15.0% (vs. 61.8%) were under the fastest half of participants to full completion of the task.

All these observations are significant with a level of at least $p < 0.01$. It should be noted that the number of marker movements did not correlate to the number of participants successfully completing the task.

Our evaluation also considered the impact of a marker as a linguistic constraint tool on communicative success. Adapted from the basic notion by Zubek (Zubek et al. 2016), we hypothesised, that the use of a marker improves team interaction as it focuses communication on critical aspects. More concretely, teams using a marker in intensive interaction should attain the highest communicative success. Based on our results this hypothesis should be rejected. We evaluate sessions with a low and a high level of interactivity in contrast with the marker condition. While the map complexity is only a control variable set from outside, the interactivity level is a phenomenon which is inherently part of team communication. Further the comparison uses the independent variables of investigation in the previous section above. In contrast, we are only interested in the degree of actors who moved the marker or wrote a comment based on a given description. We can observe that having a marker, generally becomes a disadvantage. Of all users who moved a marker once or more often, 78.6% (non-interactive sessions), 70.0% (interactive sessions) were successful in finishing the task while the marker was not visible to the describer. In contrast, when the marker was visible to the describer only 61.5% (non-interactive sessions), 40.0% (interactive sessions) finished the task successfully. Same is true if we look at the first completion time. When the marker was not visible to the describer 91.7% (non-interactive sessions), 85.0% (interactive sessions) were under the fastest half of first completion time. In contrast, when the marker was present to the describer

only 32.5% (non-interactive sessions), 29.4% (interactive sessions) were under this fast half subset. A similar observation occurs when we evaluate those sessions in which the actor applies minimally one comment. Here, we can see a disadvantage, when teams were heavily interactive and using the marker. The numbers of all users that finished with writing at least minimally one comment differs between 75.0% of teams having no marker and not being interactive to 23.1% of teams having a marker and being interactive. Furthermore, 80.0% of all teams using no marker and not being interactive ranked under half of first completion time. In contrast, interactive teams which used a marker and wrote at least one comment, achieved only 30.8% of the first completion times. With reference to marker movement only, we can further observe that interactivity was helpful to finish a task to the slowest identifier (nevertheless if it was the actor or the observer). Interactive teams where the actor moved a marker at least once, were under the fastest half of full completion times in 70.0% (no marker condition) 63.3% (marker condition) of all cases. In contrast, non-interactive teams were not so fast, only 24.1% (no marker condition) 30.8% (marker condition) were under the fastest half of full completion times.

In sessions with intense interaction while the marker was visible to the describer and all other participants, few teams completed the task (23.1%) successfully and also few subteams were under the fastest half of first completion times (30.8%). In contrast, teams with a low degree of interactivity and no marker shared with the describer achieved the best communication success. Under this condition, 75.0% of all users finished the task successfully, while 80.0% were under the fastest half of the first completion time. It can be concluded, that especially having a marker as a linguistic constraint leads to disadvantage in achieving communication success. It needs to be noted, that our first observation doesn't reach the required significance level of $p < 0.05$. Moreover, we have decided to add this observation to our considerations because the level of significance reached is very close to the required significance level and this observation supports our big picture. A further observation of the full completion time based on comments is not under consideration because it is not significant (Table 4).

8 Discussion and Conclusion

The aim of the contribution model of conversation is to reach the grounding criterion with least collaborative effort, which is fundamental to achieve communicative success based on semantic co-creation (Clark and Schaefer 1989). Linguistic constraint tools play a critical role in reaching the grounding criterion within interactions. Hanna and Brennan already mentioned the importance of constraints even outside of discussion (Hanna and Brennan 2007). Nevertheless, the applied constraint model is focused with text-processing and can not explain the adaptive nature of these constraints. The notion of linguistic constraints (Raczaszek-Leonardi and Kelso 2008; Pattee 1997) overcomes this limitation based on clear epistemological roots. Previous

Table 4 *The relationship between interaction and the marker condition on communicative success:* The independent and dependent variables are compared separately, based on the interaction level and the marker condition

Conditions	2 × 2 setting				
Interactivity level	Few interactions		Substantial interaction		
Marker condition	Not-present	Present	Not-present	Present	
Indicators of communicative success					
Degree of...	Actors, that moved the marker once or more often				p-value
All participant successfully completing the task	78.6%	61.5%	70.0%	40.0%	0.051
Fastest 50% of teams completing the task	24.1%	30.8%	70.0%	63.3%	0.003
Fastest 50% of participants completing the task	91.7%	32.5%	85.0%	29.4%	0.000
Degree of...	Actors that wrote 1 or more comments				p-value
All participants successfully completing the task	75.0%	73.3%	71.4%	23.1%	0.000
Fastest 50% of teams completing the task	35.0%	43.3%	71.4%	53.8%	0.171

research has investigated several linguistic constraint tools—like ontology (Zubek et al. 2016), contextual restrictions (Müller et al. 2013), cognitive representation model (Keilmann et al. 2017) or shared marker (Hanna and Brennan 2007). Linguistic constraint tools can help to improve team interaction and make conversation more focused (Zubek et al. 2016). Previous studies have in common that they prevent an optional use of a shared marker. The users have to move a marker to finish the given task successfully. From our point view, such a setting is unfair because it prevents the natural competition between shared marker and discussion as two parallelly applied linguistic constraint tools.

Previously, the advantage of team focused interaction was observed in complex study situations (e.g. identifying the correct wine through smelling and tasting based on a collection of available wines). Some referring expression tasks using a shared marker require a total task duration of less than 20 s. So, we asked if team focused interaction brings a general advantage, independent of the given decision complexity. Our study design compares the availability of a shared marker and two levels of decision complexity. Controlled by the number of cities in a map, we simulated two situations, where the perceived decision complexity differed. As stimuli we used a shared geographical map that contained a number of cities of which only a very small subset could be used as reference points. Furthermore, we reflected the principle of least collaborative effort as a collaborative delay discounting problem (Scherbaum et al. 2016) to the participants. The participants are not only moment-by-moment aware of how far or close they are from a state of semantic co-creation (Brennan 2005), but also the value perspective is visible to the user moment-by-moment.

Continuously the teams can evaluate the major characteristics of the contribution model of conversation. We classified sessions in low and high level of interactivity and compared them with having a marker shared with the describer or not. Based on our setup each team consists of a triad of participants (describer, actor and observer). This team configuration allows us to design a natural way of having a shared marker or not. If a marker is not shared with the describer, it still becomes useful because it is helpful between actor and observer.

We observed significant differences in the degree of team interactivity under different complexity conditions. At the lower complexity level, participants were able to use simple descriptive messages which were sufficient for accurate target identification. However, it is not enough to provide a channel for team interaction. Up to now the team focused interaction hypothesis assumes that if there are two people who can communicate it improves the identification accuracy in general. The observation in this study shows that while team interaction was possible in principle it was not really happening. One description was sent and immediately the identification happens in the next step. This pattern of discussion was already observed by Neider et al. (2010). Nevertheless, it has become obvious that team interaction by discussion itself requires a specific level of perceived decision complexity to make it an advantage. If the decision complexity is too low, real team interaction is not required.

Considerable amounts of interaction were only observed under higher complexity conditions. Nevertheless, we observed that not heavily interacting participants achieve the best performance in accuracy and task duration. That means if perceived decision complexity reaches a specific level it becomes an appropriate tool. This can be interpreted in a way that team interaction is also some kind of linguistic constraint tool, which underlies the principle of least collaborative effort. Still, we could also observe that participants interacted more intensely when they were having problems with identifying the target. Our results have shown that with a given level of complexity, a shared marker becomes useful to solve conflict situations. In similar fashion Brennan et al. (2005) already observed that if a shared marker was not present, situations evolve where in the end an actor is for a long time not far from the target and still does not identify it correctly. From our perspective a conflict situation relates to an increased perceived decision complexity that is why a marker seems to be advantageous.

In general, our results on complexity level two show, that using a shared marker was disadvantageous for the teams. In teams that had a shared marker the describer required more time to complete the task and they were less accurate in identifying the correct target. In contrast, in teams with a low degree of interactivity and no marker shared the describer achieved the best performance. It needs to be noted, that similar to our previous observation, it becomes obvious that the second identifier was faster when the team interacted heavily. From this observation we infer that the marker (as linguistic constraint tool) is not useful for the given level of complexity. Finding this certain level of complexity was not the aim of this study though.

To sum up, these observations confirm the principle of least collaborative effort towards the team focused interaction hypothesis. Team interaction and focused inter-

action are two separate linguistic constraint tools, having different characteristics, and hence need to be evaluated separately. Perceived decision complexity plays a crucial role so that team interaction becomes beneficial. Only when situations become complex team interaction is more and more helpful. Using a linguistic constraint tool (for example a marker) on top of a cognitive representation model requires a even higher level of complexity, while team interaction is already helpful. It might be the case, that if a situation is perceived as complex then in a first step we add teams and let them interact freely. If this is not enough, then we try to get interaction become more focused by using additional linguistic constraint tools.

Our observations are based on three limitations. First, the study was based on the use of a geographical map as a cognitive representation model. It is not clear whether these results can be replicated using other cognitive representation models. The characteristics of other cognitive representation models could also lead to other limitations regarding linguistic constraint tools. Additionally, using a geographical map implies a scenario in which spatial language is required to achieve a state of semantic co-creation (Spranger 2016). In other scenarios, for example searching for a target within text documents other linguistic practices of language use are required. For each different scenario appropriate linguistic constraint tools should be applied. Secondly, only a subset of linguistic constraints was evaluated. If we had only been evaluating the use of a shared marker within a shared space, we could have distinguished between different types of use, e.g. shared-eye-tracking or shared-mouse-tracking (e.g. Müller et al. 2013). A comparison of the suitability of several possible linguistic constraint tools was however not part of this study. Finally, the results show that the team focused interaction does not become true in general. A low decision complexity leads to an outstanding benefit team interaction and team focused interaction as well. The question remains open, if we generally have to reject the team focused interaction hypothesis, when we use shared cognitive representation models, or if a complex enough situation is required so that the hypothesis becomes true. Moreover, research should include the use of more complex geographical maps, containing for example 1,000 to 10,000 random locations to further evaluate the validity of this hypothesis.

Based on our results, we recommend that further studies focus on complexity when evaluating a specific cognitive representation model. The assessment of tools which are appropriate for implementation in search tasks in cognitive representation models should recommend simple filtering tools when the complexity is low. In contrast where complexity is high more sophisticated filtering tools are required to successfully complete search tasks. Using sophisticated filtering tools in complex environments leads to new situations with lower complexity where simple filters are preferred. Simple and sophisticated linguistic constraint tools can comprise any form of linguistic constraint. In our study we focused on investigating the impact of these linguistic constraints on conversation with respect to the aim of reaching the grounding criterion using the least collaborative effort.

References

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., & Sotillo, C. (1991). The hcrc map task corpus. *Language and speech,34*(4)351–366.

Brennan, S. E. (2005). How conversation is shaped by visual and spoken evidence. approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions. 95–129.

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, *106*(3), 1465–1477.

Chi, E. H. (2009). Augmented social cognition: using social web technology to enhance the ability of groups to remember, think, and reason. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data* (pp. 973–984).

Clark, H. H., & Bangerter, A. (2007). Changing ideas about reference. *Experimental Pragmatics*, 25–49.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on socially shared cognition*, 13127–13149.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science,13*(2)259–294.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*(1), 1–39.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407.

Edwards, C., & Myers, S. A. (2007). Perceived instructor credibility as a function of instructor aggressive communication. *Communication Research Reports*, *24*(1), 47–53.

Färber, M., Svetashova, Y., & Harth, A. (2020). Theories of meaning for the internet of things. *Concepts in Action: Representation, Learning and Application*, *24*(1).

Gärdenfors, P. (2004). Conceptual spaces: The geometry of thought. MIT press.

Gergen, K. J. (2009). Realities and relationships: Soundings in social construction. Harvard University Press.

Gower, J. C., Lubbe, S. G., & Roux, N. J. L. (2011). *Understanding biplots*. Wiley.

Grice, H. P., Cole, P., & Morgan, J. (1975). Logic and conversation. *Syntax and semantics*, 341–58.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*(4), 596–615.

Hanrieder, M. (2017). Collaborative spatial search: Implementation and validation of a multi-user task in walkable virtual environments.

Kaplan, R., Schuck, N. W., & Doeller, C. F. (2017). The role of mental maps in decision-making. *Trends in Neurosciences, 40*(5), 256–259.

Keilmann, F., de la Rosa, S., Cress, U., Schwan, S., Mohler, B. J., Buülthoff, H. H., & Meilinger, T. (2017). Collaborative cognition beats individual cognition: Advantages of collaborative over individual search require shared task information, and go along but also exceed the better distribution of labor. *Manuscript in preparation*.

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science, 1*(1–12), 113–114.

Kraut, R. E., Gergle, D., & Fussell, S. R. (2002). The use of visual information in shared visual spaces: Informing the development of virtual co-presence. *In Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work* (pp. 31–40).

Löbler, H. (2010). Signs and practices: coordinating service and relationships. *Journal of Business Market Management*, *4*(4), 217–230.

Logan, R. K. (2006). The extended mind model of the origin of language and culture. *In Evolutionary Epistemology, Language and Culture*, 149–167.

Monmonier, M. (2018). *How to lie with maps*. University of Chicago Press.

Müller, R., Helmert, J. R., Pannasch, S., & Velichkovsky, B. M. (2013). Gaze transfer in remote cooperation: Is it always helpful to see what your partner is attending to? *The Quarterly Journal of Experimental Psychology*, *66*(7), 1302–1316.

Neider, M. B., Chen, X., Dickinson, C. A., Brennan, S. E., & Zelinsky, G. J. (2010). Coordinating spatial referencing using shared gaze. *Psychonomic Bulletin, 17*(5), 718–724.

Pattee, H. H. (1997). *The physics of symbols and the evolution of semiotic controls*. Workshop on Control Mechanisms for Complex Systems: In Proc.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: an information search and protocol analysis. *Psychonomic Bulletin, 16*(2), 366–387.

Raczaszek-Leonardi, J., & Kelso, J. S. (2008). Reconciling symbolic and dynamic aspects of language: Toward a dynamic psycholinguistics. *New Ideas in Psychology, 26*(2), 193–207.

Sadalla, E. K., Burroughs, W. J., & Staplin, L. J. (1980). Reference points in spatial cognition. Reference points in spatial cognition. *Journal of Experimental Psychology: Human Learning and Memory, 6*(5), 516.

Scherbaum, S., Frisch, S., Leiberg, S., Lade, S. J., Goschke, T., & Dshemuchadse, M. (2016). Process dynamics in delay discounting decisions: An attractor dynamics approach. *Judgment and Decision Making, 11*(5), 472.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*(3), 379–423.

Spranger, M. (2016). The evolution of grounded spatial language. Language Science Press.

Tomasello, M. (2014). *A natural history of human thinking*. Harvard University Press.

Zubek, J., Denkiewicz, M., Dębska, A., Radkowska, A., Komorowska-Mach, J., Litwin, P..... Raczaszek-Leonardi, J. (2016). Performance of language-coordinated collective systems: A study of wine recognition and description. *Frontiers in Psychology*, 71321.