Landmarks on the move: Producing and understanding references to moving landmarks

Adriana Baltaretu, Emiel Krahmer & Alfons Maes

To cite this article: Adriana Baltaretu, Emiel Krahmer & Alfons Maes (2017) Landmarks on the move: Producing and understanding references to moving landmarks, Spatial Cognition & Computation, 17:3, 199-221, DOI: 10.1080/13875868.2016.1212863

To link to this article: http://dx.doi.org/10.1080/13875868.2016.1212863
Landmarks on the move: Producing and understanding references to moving landmarks

Adriana Baltaretu, Emiel Krahmer, and Alfons Maes

Tilburg center for Cognition and Communication, Tilburg University, Tilburg, The Netherlands

ABSTRACT

There is a general agreement that landmarks in route directions should be perceptually salient and stable objects. Yet, other attributes, such as (animated) motion, can also attract visual attention and make entities salient. In the present study, we investigate if and when speakers refer to moving entities in route directions and how listeners evaluate such instructions. We asked speakers to watch short videos of different crossroads with and without moving landmarks and give directions to listeners, who in turn had to choose a street on which to continue (Experiment 1) or choose the instruction they most preferred among three route directions (Experiment 2). Results reveal that speakers mentioned moving entities, especially when the trajectory was informative for the place where a turn should be taken (Experiment 1). Listeners had no problem understanding instructions with moving landmarks (Experiment 1). Yet, participants chose instructions with stable landmarks more often (Experiment 2). These results are discussed in relation to automatic route directions generation.

KEYWORDS

landmarks; stability; animated motion; moving landmarks; visual attention; route directions; dynamic scenes

1. Introduction

In the last decade, pedestrian navigation systems have become increasingly popular. Augmented reality, too, is slowly entering our everyday live. State-of-the-art technology can redefine the capabilities of navigation systems. For example, different devices (e.g., Tesla’s car that drives itself safely in a variety of conditions) can capture visual surroundings (with video cameras) in real time. In the future, this could enable pedestrian navigation systems to ground route directions into the visual context and generate instructions by referring to both stable database information (e.g., streets, buildings) and other type of information present in the visual field of the user. One type of information available only in the here-and-now context are moving entities (e.g., cyclists, pedestrians). We know little about how dynamic aspects of the environment can influence the production of route directions in general, and the selection of landmarks in particular. Intuitively, it is likely that, in a co-presence situation,
humans help each other with instructions such as “go left where that man turns now.” Previous research, however, hardly addressed the issue of moving landmarks, and it is still unclear if and how speakers refer to moving entities and to what extent listeners appreciate such route directions.

By route directions we refer to a set of instructions on how to (incrementally) follow a route (Richter & Klippel, 2005). In this study, we focus on references to landmarks, considered key ingredients for good route directions (Allen, 2000). Traditionally, landmarks are defined as environmental features that function as points of reference (Allen, 2000); “unique configurations of perceptual events [ . . . ] (that) identify a specific geographic location” (Siegel & White, 1975, p. 23); or, as objects that are better known and that define the location of other points (Presson & Montello, 1988). In general, in previous route direction studies, landmarks are described as concrete, route-relevant, stable entities, such as buildings.

Different scholars theorized that good reference objects are large, geometrically complex and stable (e.g., Campbell, 1993; Talmy, 1983). One likely reason for route direction studies to look upon landmarks as stable entities could be that the communicative situation typically used in the experimental setups includes some type of delay or asymmetry between producing directions and navigating with them. For example, instructions are communicated over distance (e.g., telephone) or asynchronously (on the basis of maps or previous experiences, a participant produces instructions to be used later/ evaluated by another one). In such situations, references to here-and-now events are unlikely to be produced. In this study, we focus on a situation in which this delay is absent: turn-by-turn route directions with the instructor and navigator being co-present. While experiencing a shared dynamic environment, speakers can improve the instructions by referring to anything they see.

Arguably, the most important characteristic of a landmark is its distinctiveness. Objects can be distinctive on different dimensions (for example due to familiarity or functional relevance). Here, we focus on navigation contexts that are unfamiliar to the traveller and in which perceptual salience is more important than other (e.g., knowledge-based) information. It has been theorized that the more visually noticeable or attention-grabbing an object is, relative to neighboring entities, the more likely it is to be used as a landmark (Sorrows & Hirtle, 1999). For example, color and size seem to influence landmark selection (Allen, Siegel, & Rosinski, 1978; Sorrows & Hirtle, 1999) as is confirmed by previous experimental work in natural environments (Nothegger, Winter, & Raubal, 2004; Raubal & Winter, 2002). However, in early processing stages of attention, other visual attributes come into play, such as the direction and velocity of motion (Mital, Smith, Hill, & Henderson, 2011; Treisman & Gelade, 1980). We ask to what extent this attention-grabbing property makes it likely that people refer to moving entities and if instructions with moving landmarks are preferred as much as instructions with more stable ‘traditional’ landmarks.
1.1. **Stable vs. moving objects**

In the navigation literature, it is common to think of landmark objects as being stable / permanent entities. This stability, as assumed in route direction studies, has beneficial effects on navigation and reorientation. The perceived stability of objects seems to influence toddler’s and rodent’s use of landmarks for orientation. Studies on toddler’s reorientation skills speculate that stability and scale are important factors in landmark use, where smaller or more portable objects have less navigational significance (Learmonth, Newcombe, & Huttenlocher, 2001; Smith et al., 2008). Similarly, rats can search for a location defined by visual landmarks, but will not do so if the landmark’s position has varied from trial-to-trial (for a review, see Burgess, Spiers, & Paleologou, 2004). Convergent fMRI evidence suggests that stable objects elicit greater activity in regions of the brain involved in navigation and landmark assignment (for a review, see Chan, Baumann, Bellgrove, & Mattingley, 2012).

Most of the experimental route direction studies start with the default assumption that objects have to be stable in order to be used as landmarks and with the exception of the rodent experiments, the studies mentioned above were not designed to test this assumption. In none of them is the object’s motion directly witnessed by the participants, and motion is never used as a clue that could potentially help with solving the task. Despite work on direction giving in the context of human dialogue (e.g., Brennan, Schuhmann, & Batres, 2013), as well as detailed analyses of the linguistic structure of route directions (e.g., Allen, 2000), most studies do not address situations in which both the speaker and listener have access simultaneously and can use any aspect of the visual environment.

In typical settings, speakers are asked first to learn routes from text (e.g., Ferguson & Hegarty, 1994; Lee & Tversky, 2005; Tom & Tversky, 2012), from maps (e.g., Lee, Tappe, & Klippel, 2002), by travelling the route (e.g., Denis, Mores, Gras, Gyselinck, & Daniel, 2014; Lovelace, Hegarty, & Montello, 1999; May, Ross, Bayer, & Tarkiainen, 2003; Miller & Carlson, 2011), or by free recall from memory (e.g., Denis, Pazzaglia, Comoldi, & Bertolo, 1999). Then, speakers have to provide a description and draw sketches for a fictitious listener or take a recognition test. In none of these studies do the means for acquiring and disseminating spatial knowledge allow moving entities to be mentioned. To our knowledge, studies that focus on spontaneously elicited route directions do not make any observations regarding temporarily available information (Couclelis, 1996; Denis et al., 1999; Golding, Graesser, & Hauselt, 1996; Tversky & Lee, 1998). As a result, landmarks are generally understood as stable three-dimensional objects, such as buildings and road furniture, and two-dimensional, mostly related to the path to be followed (e.g., streets) (Denis et al., 1999; Ishikawa & Nakamura, 2012; May et al., 2003).

Additionally, in the field of automatic landmark selection, buildings are considered prototypical landmarks (Sadeghian & Kantardzic, 2008), and
permanence / stability is one of the attributes considered to contribute to the perceptual salience of an object (Duckham, Winter, & Robinson, 2010; Raubal & Winter, 2002). In sum, landmarks are by default assumed to be stable, and different studies (such as the ones on orientation, route direction production and landmark selection) focus on this type of object for practical and methodological reasons.

However, it is likely that moving landmarks are part of the rich repertoire of landmark options, in particular in dynamic navigation situations in which producer and addressee are co-present. Moving entities have various features that make them potentially suitable landmarks. Motion is processed effortlessly by the visual system, and it can efficiently grab and guide attention (e.g., Abrams & Christ, 2003; Hillstrom & Yantis, 1994; Mital et al., 2011). In other words, moving entities are notably perceptually salient (Itti, 2005), a prerequisite for landmarks. In fact, motion contributes to an objects' perceptual salience as much as the combination of all other visual features (color, size, etc.) (Carmi & Itti, 2006; Itti, 2005).

Among different types of objects in motion, animate entities seem to capture attention even more. By animate entities we refer to entities conceptualized as living beings (Fraurud, 1996). Animates are conceptually highly accessible (e.g., Prat-Sala & Branigan, 2000), and visual representations of the face and the human body have the ability to capture the focus of attention even when visual attention is occupied by other tasks (for a review, see Downing, Bray, Rogers, & Childs, 2004). Humans prioritize the visual processing of animate over inanimate entities (Kirchner & Thorpe, 2006; New, Cosmides, & Tooby, 2007), and attention and animacy are linked and bias reference production (Coco & Keller, 2015).

Previous experiences may, however, hinder the use of moving landmarks, and perceptual salience might not suffice to elicit references to moving objects. People are more used to receiving and giving instructions with stable objects. After all, in everyday life, navigation systems currently do not make use of dynamic features and referring to a person that is just turning is an event that requires good timing with the direction giving process. Moreover, task demands can have a significant influence on what people watch and mention. For example, Miller and Carlson (2011) manipulated the objects’ perceptual salience (size and color) and their relevance for the navigation task (objects placed at decision versus nondecision points). Perceptual salience positively affected object memory, yet it was only the task relevance dimension that determined whether objects were included in the route directions. Congruent evidence comes from Einhäauser, Rutishauser, and Koch (2008), who showed that participants were able to suppress paying attention to moving items when searching for a different object in the scene. These results suggests that motion should be more than attention grabbing. It must be task relevant for both producer and listener. Could motion ease the instruction giving process when showing the correct path?
1.2. The current study

In route directions, people nearly always refer to landmarks (Tom & Denis, 2003), and these references maximize the helpfulness of the instructions (Allen, 2000). The type of motion perceived by speakers might influence their referential behavior. To our knowledge there are no studies that have manipulated motion as a crucial variable in their design. Do speakers refer to moving entities, and more so when they are relevant to the route task? Do listeners have problems understanding and following such instructions? To what extent do people like instructions with moving landmarks compared to stable landmarks and route directions without landmarks?

We propose investigating these questions by manipulating moving entities in a controlled, yet naturalistic environment. Participants are asked to give (Experiment 1) and choose (Experiment 2) route directions informed by dynamic scenes displaying natural everyday events. In an outdoor environment there can be different types of movement (self-produced / induced) of different entities. In this study, we focus on self-produced, animated motion.

In Experiment 1, the speaker and the listener carry out a joint direction-giving task, in which the perspectives of speaker and hearer are aligned. This contrasts with previous experimental studies that typically use an asynchronous communication setting, which imposes several spatial and temporal constraints. Based on a video, the speaker is asked to produce a route direction to a co-present addressee. Speakers can refer to any aspect of the visual environment they wish. Apart from moving entities, at each intersection, there are different stable objects available as well.

Just like ‘traditional’ landmarks, the moving entity should be placed near the location of the navigation action. In this study, there are two experimental conditions depicting moving entities that might differ in relevance for the navigation task: persons moving up to a point where they reach the intersection or taking a turn in the direction in which the listener should also turn. We do not claim that dynamic landmarks would replace the stable ones. Both moving and stable objects might be mentioned in the same utterance (e.g., “turn right at the shop, where that person is turning”). In such cases, the moving entity would help disambiguate the stable landmark with respect to a series of possible distractors (e.g., other shops in the scene).

In Experiment 2, we look at preferences for directions with different types of landmarks. Participants are tested individually in a noninteractive context. They have to choose the route direction they like best out of three options (without landmarks, with stable, or with moving landmarks). In line with previous studies that emphasize the importance of landmarks in route directions, we expect subjects to more often choose instructions with landmarks over instructions without landmarks. A research question is whether the type of movement would affect preference for the two types of landmarks.
2. Experiment 1—Production

2.1. Methods

2.1.1. Participants

One hundred and twelve native Dutch-speaking students from Tilburg University (50 women, 21.2 mean age) participated in exchange for partial course credits. They were paired in 56 dyads (i.e., groups). Participants were randomly assigned to a speaker (35 women) or a listener role. All participants gave written consent for the use of their data.

2.1.2. Materials

The materials consisted of 144 street view HD videos (108 experimental videos and 36 filler videos) recorded in Rotterdam downtown. The experimental videos depicted 36 low traffic, “+”- shaped intersections. These intersections have a simple geometric shape, in which just saying “go left,” without adding any landmark, would discriminate the target street from the other branches of the intersection. Each intersection was recorded three times illustrating a different motion manipulation as illustrated in Figure 1, corresponding to the three conditions (36 experimental videos per condition): (a) no entities moving towards / coming from the intersection (no movement condition); (b) a moving entity walking / cycling towards the intersection (irrelevant movement condition); and, (c) the same entity taking a turn in the direction required by the navigation task (relevant movement condition). Note that, since all moving entities may be to some extent relevant due to their proximity to the intersection, the terms ‘irrelevant’ and ‘relevant’ motion are used for labelling purposes only. The people recorded were casually walking / cycling down the street, without paying attention to the camera. These people were different from one intersection to another. Thirty-six filler videos were included, capturing a different set of intersections from very crowded pedestrianized areas, where individuals turning were masked by the crowd, as well as intersections with complex geometric structures in which passers-by did not turn in a direction relevant to the navigation task. These fillers were added to present participants with a range of different navigation scenarios and prevent participants from relying on fixed strategies.

2.1.3. Procedure

Participants were presented with instructions stating that the researchers were developing software that can generate real time/live pedestrian route descriptions based on the visual input coming from a Google Glass video camera and realized in audio format via a smartphone, and that at this stage they were collecting good route directions to further develop their system. Participants were given two paper booklets with line drawing maps of the intersection’s shape
(the speaker booklet included an arrow showing the direction to be taken at each intersection). The task for the speaker was to provide route directions, while the listener had to mark in his booklet the indicated street. The speaker had to first look at the map, then play the video projection and start giving instructions as soon as possible, while watching the video. The listener had to watch the video at the same time and afterwards mark the intended street on the paper map. The listener was allowed to ask questions if the instructions were unclear. Pointing was discouraged by installing a screen between participants up to shoulder level (see Figure 2). The videos were projected on a white wall, at size of approximately

Figure 1. Example of experimental clips with no movement (above), irrelevant movement (center) and relevant movement (below) in Experiment 1.
170 × 120 cm. Each video lasted about 3 seconds. The videos could not be replayed, but the last frame was displayed until the listener was finished. The video clips were divided across three lists, so that each intersection was shown only once to each participant. Participants were randomly assigned to one of the three presentation lists. The task started with 2 warm-up trials; next 72 trials (36 experimental trials) were presented in randomized order. There were no time constraints.

### 2.1.4. Design and statistical analysis

This study had Motion Type (3 levels: no motion, irrelevant motion, relevant motion) as within participants factor and Presentation List (3 levels) as between participants factor. For the first analysis, we checked if participants mentioned moving entities and if they did so more often when the motion was task-relevant. The dependent variable was coded as a binary variable: the moving entity was mentioned or not by the speaker in his or her first instruction. Next, we checked if moving entities were mentioned together with a stable object, and we analyzed listeners’ clarification questions as well as their error rates.

Statistical analyses were performed using logit mixed-model analysis (Jaeger, 2008), following the recommendations of Barr, Levy, Scheepers, and Tily (2013).
We used the mixed logit model analysis as it can correctly account for random subject and item effects in one analysis. The models were fitted using the LMER function from the LanguageR Package in R (version 2.15.2; CRAN project; The R Foundation for Statistical Computing, 2012). Motion Type and Presentation List were introduced as fixed factors and speakers and video items as random factors. The factors were centered to avoid collinearity. To determine whether the two conditions significantly differed from each other, we started by constructing a model with a full random effect structure. This maximal model included participant and video items intercepts and random slopes to account for between-subject and between-item variation. In case the model did not converge, we only excluded random slopes with the lowest variance until convergence was reached. The results from the first converging model are reported. This model included a random intercept for participants, and random intercept and random slope for Motion Type in item videos. Just as in logistic regression, the beta coefficient is a measure of how strongly each predictor variable influences the dependent variable. The p values were estimated via parametric bootstrapping over 100 iterations.

3. Results and discussion

In total, 2016 route directions (56 speakers × 36 videos) were produced. Across the three conditions, participants mentioned landmarks (N = 1113) in approximately half of the instructions of each condition (M = 0.48 in no movement; M = 0.53 in irrelevant movement; M = 0.67 in relevant movement condition). These landmarks consisted of references to stable objects (N = 752), moving entities (N = 361), and cases of stable and moving landmarks mentioned together in the same instruction (N = 43) (see Figure 3).

3.1. Moving landmarks

As expected, in the no-movement condition, participants rarely referred to moving entities. The three cases in which a moving entity was mentioned in this condition were to a person in the far distance cycling on the road that led to the intersection. Because of this, statistical analysis was performed only on data from the other two conditions, and revealed a main effect of Motion Type (β = 1.913; SE = .27; p < .001). In the relevant movement condition participants referred more often to the moving person taking a turn (M = .37, SD = 0.48), than in the irrelevant movement condition (M = .13, SD = 0.34). There was no significant effect of Presentation List (p > .05) and no interaction between the two factors (p > .05). Some examples of instructions with moving landmarks are given next. (English translations of Dutch originals). Speakers focused either on the location of the moving entity (examples 2 and 4) or on the action that the man was doing (example 3).
Referring to the pedestrian’s location seemed to be the most frequently used strategy in this corpus.

1. You need to turn left, on the same side where the man is walking.
2. Go right, where the man is going.
3. Go on the first street right, follow the cyclist.
4. You need to go left, where the pedestrian is.

### 3.2. Moving and stable landmark combinations

We analyzed if moving landmarks were mentioned in the same instruction together with stable ones. In $N = 361$ route directions in which the moving person was mentioned, 11% cases included both moving and stable landmarks ($N = 1$ in the no movement condition, $N = 20$ in the irrelevant movement condition, $N = 22$ in the relevant movement condition, see also Figure 3). The order in which the mixed landmarks were introduced in the instruction was mostly consistent, with moving landmarks being mentioned before the stable ones 80% of the times. Following are some examples of mixed landmarks instructions from the irrelevant movement condition (examples 1 and 2) and the relevant movement condition (examples 3 and 4) for which stable and moving landmarks are italicized:

1. Where the cyclist is, go first [street] right, after the sign with the lion.
2. Can you see where the girl is walking? You need to go between KFC and Febo.
3. Go right, first on the right at the man, before the car.
4. Go straight until the stairs and then turn left where the people are also walking.

### 3.3. Clarification questions and error rates

Listeners were allowed to ask clarification questions. Their questions could indicate whether or not instructions with moving landmarks were harder to understand and follow.

Sometimes, listeners repeated parts of the route directions while drawing, uttered confirmatory remarks, or told the speaker to continue with the next trial. In general, the task was easy, there were few questions and there were no signals of major communication breakdowns in which the speaker and the listener failed to understand each other. Listeners’ questions were analyzed using the Stivers and Enfield (2010)’s coding scheme for question-response sequences in conversation. According to this scheme an utterance was considered a question if it was a formal question (it had lexico-morpho-syntactic or prosodic interrogative marking) or a functional question (it had to elicit information, confirmation or agreement). Apart from 11 utterances that could not be evaluated due to technical issues, there were 81 questions, posed by 30 listeners. These questions were asked when the speaker did not include any landmark in his initial instruction (49%), when the speaker referred to a stable landmark (30%), a moving landmark (19%) or to both stable and moving landmarks in the same instruction (2%). Compared to the other conditions, there were more questions posed in the no-movement condition (44% in no-movement, 26% in irrelevant movement, 30% in relevant movement condition). The questions were classified on the basis of the semantic structure of the utterance, and next the question’s goal (or social action in Stivers and Enfield’s terms) was assessed.

Based on the logical semantic structure of the utterance, questions were classified as polar yes/no question (a question that elicits a confirmation/disconfirmation), alternative questions (questions that included a restricted set of alternative answers), or content (mostly questions seeking information, introduced by a ‘WH-word,’ such as what, where, etc.). The frequency with which these types of questions occurred, the goal for which they were used, and examples are given in Table 1.

These questions might represent problems of various levels of difficulty, and one might expect more content related questions if the task would be difficult. Yet, most of the questions (60 cases) were simple polar questions, asking for confirmation. About half of these (33 cases) were asked after the speaker did not include any landmark in his first instruction. Listeners asked for confirmations with respect to three aspects. There were questions related to the direction of the turn (e.g., ‘left?’, 16 cases), to correctly choosing the street
(e.g., ‘first [street]’?, 16 cases) and questions about the place where to turn (22 cases). When asking about the place to turn, listeners added references to stable (e.g., ‘turn after the shop’?, 19 cases) and moving landmarks (e.g., ‘turn / go after that man’?, 3 cases) or mentioned the path (e.g., ‘in the first intersection’?, 8 cases).

Second, requests for clarification (12 cases) resulted from minor misunderstandings, although the speakers’ instructions generally had landmarks (10 out of 13 instructions had landmarks). Listeners asked for direction clarifications (8 out of 12 cases, e.g., ‘left or right?’), and in five of these questions stable landmarks were mentioned. The rest of the questions were about the manner of movement through the intersection (e.g., ‘am I on the right or on the left side of the street?’). These might be related to the listener’s task, namely drawing the route on paper maps, and listeners wanted to know in more detail how to move in intersections.

Third, there were few requests for information. There were six ‘WH’ questions, one including a moving landmark reference. In sum, listeners posed questions primarily to confirm their choice, and to a lesser extent, to find out more information about the street they should follow.

Finally, listeners made few errors when choosing the street indicated by speakers. There were only 11 cases in which they marked a wrong street on their maps and another 8 cases in which they corrected their initial choice. The small number of errors suggests that the task was simple and listeners had no real problems accomplishing it.

To sum up, these results revealed that speakers mentioned moving entities when giving route directions. They especially did so when the movement trajectory was informative for the place where a turn should be taken. Moving entities were rarely mentioned together with stable ones, which might indicate that moving entities can be seen as an alternative type of landmark. Listeners
had no difficulty in understanding and using instructions with moving landmarks. They rarely asked for clarification and made few errors choosing the correct street, after hearing an instruction with a reference to a moving target. In particular, there was a similar number of clarification questions when speakers referred to stable and to moving landmarks, which makes it interesting to look in more detail at participants’ preferences for these entities. To do so, in Experiment 2, we showed videos depicting irrelevant and relevant movement and asked participants to choose the instruction they prefer most.

4. Experiment 2—Evaluation

4.1. Methods

4.1.1. Participants

Thirty-two native Dutch-speaking students of Tilburg University (12 women, 20.7 mean age) participated in exchange for partial course credits. None of them participated in Experiment 1. All participants gave written consent to the use of their data.

4.1.2. Materials

The materials consisted of 72 videos (the experimental trials from the irrelevant and relevant movement conditions) used in Experiment 1. Overlaid on the videos, a semitransparent red arrow depicted the route and the direction to be followed (see Figure 4).

For each video, we created a set of three route directions as follows: First, we analyzed the data coming from 10 speakers from Experiment 1 that referred most often to landmarks (irrespective of condition and type of landmark mentioned). The directions could have different information structure (see Experiment 1 for examples). The most frequent word order they used (about 80% of the cases) consisted of a verb and the direction of turn, followed by a landmark. Next, starting from this dominant template, we created for each video three different route directions: one without landmarks, consisting only of a motion verb and the direction (e.g., “turn left”); a route direction with a stable landmark (e.g., “turn left at the Hema shop”) and a route direction with a moving landmark (e.g., “turn left where that man / woman / cyclist is going”).

In Experiment 2, all the instructions had the same information structure as the one most frequently observed in Experiment 1, namely a verb + direction + stable (e.g., “at the pharmacy”) or moving landmarks (e.g., “where the [man / woman / cyclist] is going”), in this exact order. The stable landmarks used in these route directions were the most often mentioned stable objects in Experiment 1. For the moving landmarks, we used the most frequent referring expressions (the man / the woman / the cyclist).
As in Experiment 1, participants were presented with instructions stating that the researchers were developing software that can generate real time/live pedestrian route descriptions based on the visual input coming from a Google Glass video camera and realized in audio format via a smartphone. The participants’ task was to evaluate different route directions produced by this application. Participants had to watch a number of videos, and for each read the three route directions and choose the one that they liked most. The videos could not be replayed, but the last frame of the video was visible until the participants clicked on the route direction of their choice. Two presentation lists were created, so that each intersection was shown only once to each participant. The trials consisted of a fixation cross-displayed for 500 ms, followed by a slide that displayed the video on the upper three quarters of the screen, and three instructions placed below the video, next to each other. The position on screen of the three types of route directions was counterbalanced and randomized, so that each type would be displayed an equal amount of times on the left, center and right side of the screen. Once the choice was recorded a new trial would automatically start. The experiment began with two

![Figure 4](image1.png)

**Figure 4.** Example of experimental clips with irrelevant movement (above) and relevant movement (below) with arrows showing the turning direction in Experiment 2.

### 4.1.3. Procedure

As in Experiment 1, participants were presented with instructions stating that the researchers were developing software that can generate real time/live pedestrian route descriptions based on the visual input coming from a Google Glass video camera and realized in audio format via a smartphone. The participants’ task was to evaluate different route directions produced by this application. Participants had to watch a number of videos, and for each read the three route directions and choose the one that they liked most. The videos could not be replayed, but the last frame of the video was visible until the participants clicked on the route direction of their choice. Two presentation lists were created, so that each intersection was shown only once to each participant. The trials consisted of a fixation cross-displayed for 500 ms, followed by a slide that displayed the video on the upper three quarters of the screen, and three instructions placed below the video, next to each other. The position on screen of the three types of route directions was counterbalanced and randomized, so that each type would be displayed an equal amount of times on the left, center and right side of the screen. Once the choice was recorded a new trial would automatically start. The experiment began with two
practice trials, followed by 36 experimental trials presented in random order. There were no time constraints.

4.1.4. Design and statistical analysis

This study had Motion Type (2 levels: irrelevant motion, relevant motion) as within participants factor and Presentation List (2 levels) as between participants factor. The dependent variable was the type of route direction chosen. Statistical analysis was performed as in Experiment 1. Motion Type and Presentation List were included as fixed factors; subjects and videos as random factors; random intercepts and random slopes for speakers and videos. The first converging model is reported (p values were estimated via parametric bootstrapping over 100 iterations). This model included random intercept for subjects and random intercept for videos.

5. Results and discussion

Out of 1152 cases (36 scenes \( \times \) 32 participants), route directions with landmarks were chosen more often (73% of the cases) than route directions without landmarks (see Figure 5). To check if motion influenced the choice for a specific type of landmark, the statistical analysis was done on a data set consisting of only the route directions with landmarks.

There was a main effect of Motion Type (\( \beta = 1.211; SE = .265; p < .001 \)). For videos depicting irrelevant movement, participants chose more often instructions with stable landmarks (\( M = .85 \)) than with moving landmarks (\( M = .15 \)). For videos depicting relevant movement, the same pattern is

![Figure 5](image-url)  
*Figure 5.* Number of route directions types chosen across conditions in Experiment 2.
observed (stable landmarks $M = .75$; moving landmarks $M = .25$). There was no significant effect of Presentation List ($p > .05$), and no interaction between the two factors ($p > .05$).

There was a slight increase in the preference for moving landmarks in the relevant movement condition. To test if this observation was statistically significant we ran a Wilcoxon matched-pairs test. There was a significant difference in the scores for irrelevant movement ($M = .15$) and relevant movement ($M = .25$), $z = -2.67, p < .05, r = -.65$.

In conclusion, participants preferred instructions with landmarks over instructions without landmarks. Stable landmarks were chosen more often than moving landmarks, as the preferred ones. Compared to the ease with which listeners used moving landmarks in Experiment 1, the overwhelming preference for stable landmarks is striking.

6. General discussion

In this study, we manipulated the trajectory of moving entities walking in intersections and analyzed if and when speakers refer to moving entities (Experiment 1) and participants’ preferences for these references (Experiment 2). In the first experiment, we used an interactive setting, in which speakers were giving instructions to listeners, the two having the same visual information and access to the same navigation environment (videos of intersections). Speakers included (both types of) landmarks in approximately half of the instructions. Human speakers do refer to moving landmarks in the co-presence situation. Motion is perceptually salient, however not only salience, but also task relevance greatly contributed to our results (see also Miller & Carlson, 2011). Moving landmarks were mentioned especially in the relevant motion condition. In fact, in this condition, there were more references to moving than to stable landmarks. Quite often the moving entities were mentioned alone without further references to stable objects, suggesting that moving items can be used as landmarks on their own. Listeners did not encounter problems understanding such instructions, as revealed by the analysis of their questions and error rates. They (rarely) asked for clarifications, and the questions they did ask were mostly due to minor misunderstandings.

In the second experiment individual participants had to choose an instruction for videos depicting irrelevant and relevant motion. Instructions with landmarks were preferred over instructions without landmarks. This result is in line with previous studies that underline the importance of landmarks at decision points where reorientation is required (Denis et al., 1999). In addition, route directions with stable landmarks were preferred over instructions with moving landmarks. Next, these results are further discussed in relation to route direction production, and possible suggestions for automatic route direction generation are proposed.
6.1. Moving landmarks in route directions

In earlier research, landmarks are considered stable entities, and there are indeed many communicative situations where stable objects are valid, relevant landmarks (route directions from maps, memory, etc.). The asynchronous situations used in previous studies favor an approach that defines landmarks as objects that are stable, memorable and stand out in the general knowledge one has about the environment. Yet, moving entities can be useful landmarks under specific circumstances (direction giving as a joint activity and while observing moving entities during utterance planning). They seem to be effective in route directions due to their perceptual salience, but more so when informative. In our setup, moving entities are relevant because they show the right turn in the intersection. Because motion involves both spatial and temporal dimensions, the main role of these moving entities was to provide short-term orientation for the listener and locate with their presence the street where a turn should be made. In addition, the joint perception of an entity moving on a relevant trajectory affords the visualization of an emerging action: the future turning action of the listener. An instruction such as ‘follow the man in red’ can subsume several pieces of information. It tells the listener the place where to turn, and the direction in which the turn should be done, and it singles out the street on which to continue.

To evaluate how listeners cope with such instructions, we analyzed the listeners’ clarification questions and their error rates. In general, there were few errors and the task was simple. Listeners rarely asked clarification questions and their drawings were mostly correct. Interestingly, most of the questions were asked when the speaker’s instruction did not contain landmarks. Listeners introduced both moving and stable landmarks in their questions, however stable landmarks predominated. There could be different explanations for this choice, such as a possible listener’s preference for stable landmarks or an effect of the timing of the events: by the time listeners formulated their questions, the entity would have moved out of sight.

To further test if moving landmarks were considered as good as stable ones, we asked a different group of participants to watch videos depicting relevant or irrelevant motion and choose the instructions they liked best. They were presented with three instructions: without landmarks, with stable or with moving landmarks. Participants preferred instructions with landmarks over instructions without landmarks. Stable landmarks were chosen more often than moving landmarks. When motion was task-relevant, the preference for moving landmarks slightly increased. Yet, the increase in preference for moving landmarks was rather small. Given that listeners can adapt and use moving landmarks fairly easily, the preference for stable landmarks might have been influenced by the noninteractive context in which the task took
place. Another possible explanation could be that there is a speaker-listener asymmetry. Movement might primarily capture speaker’s attention, and speakers do not necessarily tailor their utterance to the listener’s needs. Further research is needed to evaluate a possible effect of the joint communicative situation on landmark preference.

Overall, the data collected in the two experiments suggest that landmark references were influenced by dynamic events present in the environment. However, one could expect that the frequency with which moving landmarks are mentioned depends heavily on the scene. The scenes studied in these experiments are admittedly relatively simple. For example, there was only one moving entity present in each scene. Real environments are noisier, and we can assume that traffic and visual clutter (e.g., surrounding pedestrians and cars) would influence the choice of landmark objects. An interesting question for future research is how more complexity (e.g., multiple moving entities, complex intersections) affects the usability of moving landmarks.

If an entity moves in the relevant direction, the speaker would have to minimally say “follow it,” avoiding the burden of describing the turn or the street. This might ease the speaker’s task, especially when describing complex turning actions (Klippel, Tenbrink, & Montello, 2013). For the addressee, following another person may be beneficial, as well (Krieg-Briickner, 1998, for ‘following’ as a basic navigation strategy). Moving landmarks bear some similarity to linear landmarks (Richter & Klippel, 2004), which afford being followed (e.g., “follow the river”). Yet, a moving entity can only guide in one specific direction (unlike linear landmarks, such as rivers, which stretch in two directions), and they are not suitable for determining actions in several decision points (e.g., follow that person for three intersections).

Any navigation help used in route directions might be considered a landmark (Couclelis, Golledge, Gale, & Tobler, 1987; Richter & Winter, 2014). Landmarks could range from personal events, as long as the latter are salient in memory and part of our shared knowledge (e.g., I tripped on the curb, and I broke my ankle, and you know this place), to concrete, stable objects that have been discussed in the literature. In between, there could be other types of landmarks, such as intersections, linear (rivers) and area-like (forests) objects. A moving landmark might be placed somewhere between these classes: it can attract attention like a classical landmark, but it is in some sense personal, because it is an ephemeral event witnessed by a small group of people. Experimental study of these aspects is greatly needed before having a more clear position regarding the role of moving entities in broader theories on navigation.

Finally, it is worth investigating if the effect observed is due only to movement being relevant or also to animacy. Would speakers still mention the moving landmark if this would be a car? We suspect it is movement and showing the right turning direction that matters.
6.2. **Implication for automatic route directions generation**

The automatic generation of route descriptions is often studied within Natural Language Generation, NLG (e.g., Dale, Geldof, & Prost, 2005; Roth & Frank, 2009). NLG systems typically involve Referring Expression Generation (REG) (e.g., Krahmer & Van Deemter, 2012) for generating references to landmarks. Until recently, REG for landmarks and studies of route direction production have focused exclusively on references to stable entities. Given the results of this study we will sketch some suggestions regarding automatic generation of moving landmark references.

Though there is room for improvement, human detection and tracking in videos has reached a quality level that would allow navigation systems to make use of dynamic aspects from the environment (Dollar, Wojek, Schiele, & Perona, 2012; Venugopalan et al., 2015). For generating more human-like instructions, navigation systems could include references to moving objects in route directions, as long as estimations of the eye gaze are available to track what the listener is watching and the instructions are well timed with the event. Our data shows that speakers spontaneously refer to items with a task relevant trajectory. Incorporating references to dynamic landmarks could optimize route directions, by providing a wider range of possibilities for choosing landmarks. For example, buildings might not be present or are hard to refer to. Sometimes buildings do not present any particularity (such as shops or architectural features) and might have indefinite colors or a small size contrast with other buildings. In such situations, moving landmarks seem a natural option and listeners should successfully handle such instructions.

Two aspects pose further challenges for integrating this type of reference in route directions. First, our data showed that, in a noninteractive context, participants preferred stable landmarks. It is an empirical question if users engaged in dialogue with a system would consider moving landmarks as good as stable ones. The efficiency of these instructions during navigation needs to be assessed. Second, little is known about the relation between motion and reference production and comprehension, and this poses specific challenges for REG.

Typically, REG algorithms produce references by using a predefined list of preferred attributes. An object is first referred using a noun (e.g., go left at the building). If this description does not single out the object from the scene (there are several buildings), more attributes (e.g., color, size) are added until all the other objects are eliminated. Some of these attributes are used more often (are preferred) over others (e.g., color is often mentioned, unlike size or location) and algorithms would typically make use of this preference (Santos Silva & Paraboni, 2015; Krahmer & Van Deemter, 2012). However, motion has been shown to attract visual attention even more than color, intensity, or orientation features (Carmi & Itti, 2006; Itti, 2005; Mital et al., 2011). More experimental
work is needed to investigate how motion relates to these visual cues. Many studies that examined static scenes provided valuable insights on how perceptual saliency affects attribute preference (e.g., Clarke, Elsner, & Rohde, 2013; Viethen & Dale, 2008), yet the scalability of their conclusions to the dynamic world remains an open question.

Finally, cutting edge technology provides machines with a rich sensory input. Our results suggest that movement detected in the nearby environment can be informative in a landmark selection task. These results highlight that not only the stable landmarks, but the moving entities, play a role in the production of turn-by-turn route directions.

**Funding**

The first author received financial support from the Netherlands Organization for Scientific Research, via NWO Promoties in de Geesteswetenschappen (322-89-008), which is greatly acknowledged. Partial results of this study have been presented in CogSci 2015 and ENLG 2015.

**References**


Klippel, A., Tenbrink, T., & Montello, D. R. (2013). The role of structure and function in the conceptualization of directions. In M. Vulchanova & E. van der Zee (Eds.), *Motion encoding in language and space* (pp. 102 – 199). Oxford: Oxford University Press.


