Stories in the Mind?

The Role of Story-Based Categorizations in Motion Classification

Frank Papenmeier[1], Juan Purcalla Arrufi[2], & Alexandra Kirsch[3]

[1]Department of Psychology, University of Tübingen, Germany

[2]Department of Computer Science, University of Tübingen, Germany

[3]Independent Scientist, Tübingen, Germany

This is a preprint of our manuscript, last date of modification: 21 June 2022

Keywords: spatial representation, qualitative reasoning, perception, short-term memory

Author Note

Correspondence concerning this article should be addressed to Frank Papenmeier, University of Tübingen, Schleichstr. 4, D-72076 Tübingen, Germany. E-mail: frank.papenmeier@uni-tuebingen.de

Word count: 8237 (main body of text); Abstract: 220

Abstract

Categorization is a central concept for spatial representations in human cognition and artificial intelligence. With the present research, we aimed at building a bridge between those two research fields by asking whether motion categorizations designed in artificial intelligence can inform the psychological understanding of human perception and memory of motion scenes. We report the results of four experiments investigating the influence of Motion-RCC and Motion-OPRA$_1$ categorizations on human perception and memory. Participants viewed simple motion scenes and judged the similarity of transformed scenes with this reference scene. Those transformed scenes differed in none, one or both Motion-RCC and Motion-OPRA$_1$ categories. Importantly, we applied an equal absolute metric change to those transformed scenes. When the reference stimulus and transformed stimuli were visible at the same time (Experiments 1a and 1b: perception), both Motion-OPRA$_1$ and Motion-RCC influenced the similarity judgements, with a stronger influence of Motion-OPRA$_1$ than Motion-RCC. When the participants first memorized the reference stimulus and viewed the transformed stimuli after a short blank (Experiments 2a and 2b: memory), only Motion-OPRA$_1$ had marked influences on the similarity judgements. We conclude that human perception and short-term memory utilize some of the properties underlying those categorizations and given this link between human cognition and qualitative reasoning, we argue for a continued and close multidisciplinary approach to investigating the spatial representation of motion scenes.

1. Introduction

Categorization lies at the heart of how humans perceive the world (Lakoff, 1987; Murphy, 2002), and it is “the most basic phenomenon in cognition” (Cohen & Lefebvre, 2005, p. 2). Categorization research has traditionally focused on objects (Medin & Heit, 1999), but it also concerns actions and events (e.g., Zacks & Tversky, 2001). In the present paper, we focus on categorizing situations that involve the movement of two entities, which is similar to situations that might occur in walking behavior (Basili et al., 2013) or traffic.

Spatial cognition is one of the few fields with extensive interdisciplinary interaction between the descriptive approach of psychology and neuroscience, and the prescriptive approach of engineering disciplines such as robotics and artificial intelligence. However, the specific subfield of qualitative spatial representations lacks psychological studies testing cognitive plausibility (Renz et al., 2000; Yang et al., 2015). With the present work, we contribute to reducing this research deficit. We studied whether categorical representations originally designed in the context of qualitative reasoning methods in artificial intelligence might help inform the representation of dynamic scenes in human perception and memory.

In the following, we first give an overview of the role of categories and inter-object relations on spatial representation in human perception and memory. After that, we introduce the concept of the qualitative representation of motion scenes from the perspective of artificial intelligence. Finally, we report four experiments that we performed to study the influence of those qualitative representations on human perception and memory.

1.1. Spatial Structure in Human Perception and Memory

A group of objects is not perceived and represented as multiple individual objects in isolation, but those objects are organized into meaningful structures – this is a core principle underlying both human perception and memory. The following will give a short overview of multiple research areas from psychology that highlight the role of spatial inter-object relations in perception and memory. Because our primary interest considers motion stimuli, this overview focuses on spatiotemporal properties.

The Gestalt principle of common fate (Wertheimer, 1923) is one of the first formal descriptions of the perceptual grouping of objects based on spatiotemporal properties. According to this principle, objects undergoing a common change (e.g., moving together) are perceptually grouped. This common motion does not necessarily require objects to move in the same direction, but grouping can also occur with objects undergoing other forms of coordinated movement, such as objects moving along a circular path. Thus, human observers can use motion cues to group or categorize objects within dynamic scenes.

Further evidence that human observers process multiple moving objects in relation to one another comes from research using the Multiple Object Tracking paradigm (MOT; Meyerhoff et al., 2017; Pylyshyn & Storm, 1988). Within this paradigm, observers try to track a subset of moving target objects among identically looking moving distractor objects with visual attention. Yantis (1992) found that observers' tracking performance was higher for conditions constraining target objects' motion to remain convex compared to conditions allowing for target motion that resulted in a collapse of the virtual polygon formed by the targets. Further, studies recording eye movements during MOT found that participants direct their gaze not only toward the individual target objects but also toward the centroid of the virtual polygon formed by the target objects (Fehd & Seiffert, 2008, 2010; Huff et al., 2010; Zelinsky & Neider, 2008). Thus, observers track objects not in isolation but in relation to one another.

Short-term memory underlies tight capacity limitations (Cowan, 2001), and the organization of information in short-term memory is a core question tackled by prior research (e.g., Jiang et al., 2000; Wood, 2011). Regarding the representation of spatial information, previous research found that objects are represented based on the spatial configuration formed by the memorized objects (Hollingworth, 2007; Jiang et al., 2000; Papenmeier & Huff, 2014; Timm & Papenmeier, 2019). For example, Jiang et al. (2000) asked participants to memorize the locations of multiple concurrently presented objects individually. Despite this instruction, and although participants were required to detect the location change of one individual object highlighted during retrieval, observers still memorized the spatial relation of the objects. That is, observers' ability to detect location changes was much higher when the probed object was presented together with the spatial configuration of all encoded objects rather than when it was presented either alone or within a distorted spatial configuration during retrieval. Within the domain of action control, studies reported similar findings. That is, reaching actions toward the location of a memorized target object were biased by shifts of the spatial configuration of other potential reaching targets (Klinghammer et al., 2015, 2016, 2017). Thus, spatial configurations underly both short-term memory and action control.

Beyond spatial information, also spatiotemporal information is represented based on STM configurations (Papenmeier et al., 2012; Sun et al., 2015). Whereas Papenmeier et al. (2012) provided first evidence that the representation of objects in dynamic scenes is based on the global spatial configuration formed by the moving objects, Sun et al. (2015) provided a more fine-grained analysis: they studied the maintenance of dynamic spatial configurations in memory. In their experiments, participants watched four moving dots and they manipulated the geometric properties of the polygon resulting from the dots. Whereas dots moving such that the polygon converted from a convex polygon into a concave polygon broke the dynamic spatial configuration, other geometrical transformations that retained the polygon as convex did not. This suggests that there are geometric categories that determine dynamic spatial configurations to some extent. However, in this previous research, metric changes were not controlled for.

1.1.1. Categorical and Metric Representations

Spatial relations between objects can be described both in categorical terms, such as *above/below* or *near/far*, as well as metric terms, such as *distance* or *angle*. The brain subsystem associated with categorical coding is located in the left hemisphere and the subsystem associated with metric coding is located in the right hemisphere (Kosslyn, 1987). However, this relationship is also mediated by task demands and practice (Banich & Federmeier, 1999).

It is typically assumed that humans represent both categorical and metric spatial information concurrently and that categorical information can be used, for example, to adjust inexactly represented metric information (Huttenlocher et al., 1991, 2000). That is, the separate representations of categorical and metric information are combined when trying to relocate the spatial location of objects (Sampaio & Wang, 2017). When participants tried to relocate a dot that was presented within a disc, for example, they reported a dot location that was biased toward the diagonal of the respective quadrant of the disc the dot was located within (Huttenlocher et al., 1991; Sampaio & Wang, 2010, 2017). In a similar vein, participants judging multiple stimuli along simple dimensions, such as the length of a lines or the fatness of fishes, they spontaneously construct categories reflecting the range of presented stimuli with individual judgements being biased toward the mean of those categories (Huttenlocher et al., 2000).

To summarize, researchers found evidence that categorical representations bias metric responses for individual objects (Huttenlocher et al., 1991, 2000; Sampaio & Wang, 2017) and that inter-object relations are used to represent both spatial and spatiotemporal information in perception and memory (Jiang et al., 2000; Papenmeier et al., 2012; Sun et al., 2015; Wertheimer, 1923; Yantis, 1992). However, we lack research on what categorical information observers might use to represent dynamic scenes in perception and memory. In the present research, we intend to close this gap by combining findings from psychology with ideas on spatial representations from robotics and artificial intelligence reported below. More specifically, by using simple motion scenes depicting two moving objects, we investigated the role of categorical properties derived from qualitative representations of motion on human perception and memory.

1.2. Qualitative Representations in Artificial Intelligence

Spatial relationships in static situations have been modeled for a long time in the field of qualitative representation and reasoning (Dylla et al., 2017). Qualitative representations describe spatial situations in meaningful categories rather than metric sensor data, such as coordinates; and, at the same time, simplify the profusion of numeric data into a few categorical values. Two main classes of qualitative spatial representations of static scenes have been proposed and extensively researched in artificial intelligence (e.g., Gantner et al., 2008; Lücke et al., 2011): (1) *Directional representations* such as one object being “to the left of” another object or one object being “in front of” another object and (2) *regional representations* such as one object overlapping another object or one object being enclosed by another object.

OPRA$_m$ (Oriented Point Relation Algebra; Mossakowski & Moratz, 2012) is a class of directional representations, where *m* represents the granularity of the division. In the present work, we have used OPRA$_1$, which divides the space into four regions through a single line. These regions are left of the line, right of the line, forward, and backward, and they are named as '0', '1', '2', '3' (see Figure 1A). Within this representation, all objects are assumed to be points in space, that is, the size of the objects is disregarded. In OPRA$_m$, the spatial relation between two entities is expressed as $\angle^{a}_{b}$, where *a* and b are values of the regions (0,1, 2, ...) of each entity with respect to the other entity, that is $\angle^{k\ with\ respect\ to\ l}_{l\ with\ respect\ to\ k}$ (see Figure 1B for an example).

(A) (B)

Figure 1. (A) Example of OPRA$_1$ dividing the surrounding space into the regions 0, 1, 2, and 3. (B) Spatial relation $\angle^{0}_{3}$ between two objects k and l as represented in OPRA$_1$.

Regarding the regional representation of two objects, the relations of RCC (Region Connection Calculus; Randell et al., 1992) are a typical representative. Figure 2 shows the RCC spatial relations. In our experiment, one entity was always smaller than the other entity; therefore, only spatial relations DC, EC, PO, TPP, and NTPP were possible (see Figure 2).

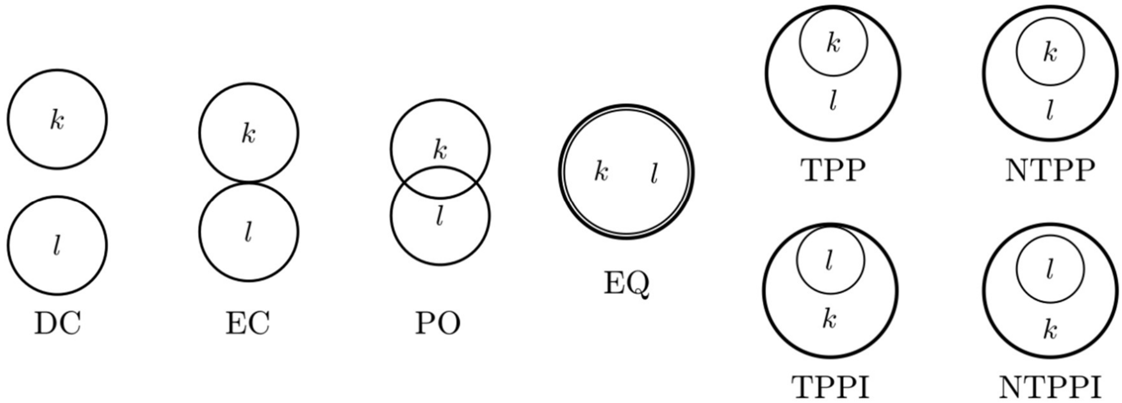


Figure 2. RCC spatial qualitative representation for rigid regions with the following relations: DC: disconnected, EC: externally connected, EQ: equal, PO: partially overlapping, TPP: tangential proper part, TPPi: tangential proper part inverse, NTPP: non-tangential proper part, NTPPi: non-tangential proper part inverse. Note that the relations EQ, TPP, NTPP, TPPi, and NTPPi depend on the relative size of the entities.

1.2.1. Story-based Representations of Motion Scenes

Recently, Purcalla Arrufi and Kirsch (2018) introduced the concept of *stories* that allow for categorizing spatial relations in motion scenes. Stories are temporal sequences of spatial relations. For example, if one watches two objects moving toward another, then touching another, and moving away from another, this could be represented as the story disconnected (DC), then externally connected (EC), then disconnected (DC) based on RCC. In a story, the order of the spatial relations implicitly represents the temporal dimension of the motion scene. Purcalla Arrufi and Kirsch (2018) applied the concept of stories to two classes of qualitative spatial representations, namely $OPRA_1$ and RCC. Based on these spatial relations, they derived the story-based categorizations of Motion-$OPRA_1$ and Motion-RCC, respectively. Figure 3 shows an example of how Motion-$OPRA_1$ and Motion-RCC can represent the same motion scene. Importantly, similar to $OPRA_1$ and RCC providing a parsimonious spatial representation of static scenes, the resulting number of possible stories for Motion-$OPRA_1$ and Motion-RCC is finite and relatively small, such that those stories can be used as categorizations and parsimonious representations of motion scenes.

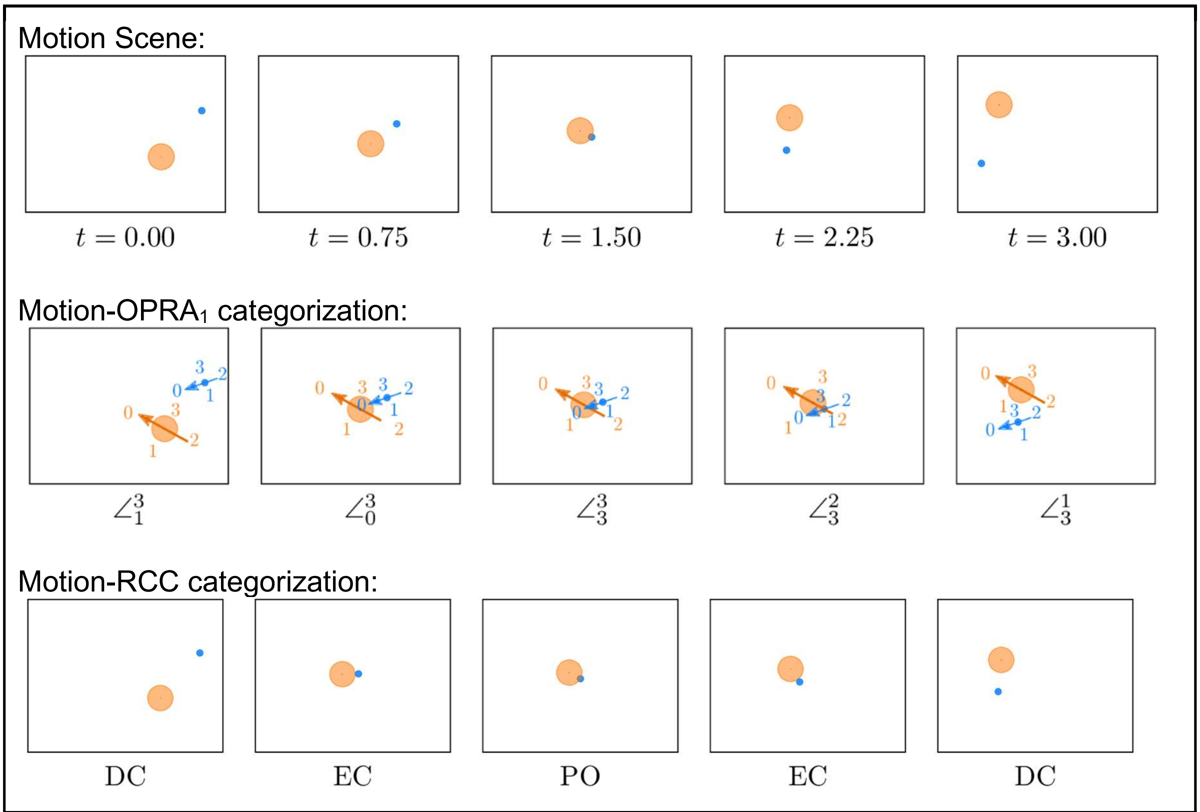


Figure 3. Illustration of how Motion-OPRA$_1$ and Motion-RCC categorize *the same motion scene*

- The snapshots at the top row represent the complete motion scene (from t=0.0 to t=3.0 s) as snapshots that are equally spaced in time.
- The snapshots in the middle row – not equally spaced in time – show all the $\angle^3_1$ $\angle^3_0$ $\angle^3_3$ $\angle^2_3$ $\angle^1_3$ successive different OPRA$_1$ spatial relations, which we interpret below:
    - $\angle^3_1$ the small blue disc is on the right (3) of the large orange disc, the large orange disc is on the left (1) of the small blue disc
    - $\angle^3_0$ the large orange disc is directly in front of the small blue disc (0), while the small blue disc remains on the right (3) of the large orange disc
    - $\angle^3_3$ the center point of the large orange disc is on the right (3) of the center of the small blue disc, the small blue disc is still on the right (3) of the large orange disc
    - $\angle^2_3$ the small blue disc is directly behind (2) the large orange disc, the large orange disc is on the right (3) of the small blue disc
    - $\angle^1_3$ the small blue disc is on the left (1) of the large orange disc, the large orange disc is on the right of the small blue disc (3).

  We abbreviated this Motion-OPRA$_1$ story as mo-on.
- The snapshots at the bottom show all the successive different RCC: the entities approach (DC), touch (EC), partially overlap (PO), touch again (EC) and move apart (DC). We abbreviated this Motion-RCC story as mr-po.

1.3. Present Research

So, is there a relation between the story-based representations of motion scenes developed in artificial intelligence and the way humans represent motion scenes during perception or memory? With our present research, we aimed to tackle this question by asking whether the categorizations defined in Motion-$OPRA_1$ or Motion-RCC might reflect aspects that humans use to represent motion scenes during perception or memory. Thus, we ran four experiments asking participants to view or memorize simple motion scenes (reference scene) and to determine the similarity of transformed scenes with this reference scene. Critically, all transformed scenes consisted of metric changes of the same absolute amount as compared to the reference scene, but they differed in whether they were of the same or a different Motion-$OPRA_1$ or Motion-RCC category. If the Motion-$OPRA_1$ or Motion-RCC categories are meaningful to human perception or memory, a change in a respective category should result in reduced similarity judgements. Otherwise, the similarity ratings should be unaffected by changes in the respective Motion-$OPRA_1$ or Motion-RCC categories.

2. Experimental Set 1: Perception

We performed two experiments (Experiment 1a and Experiment 1b) in order to determine the influence of Motion-RCC and Motion-$OPRA_1$ story-based categorizations on the human perception of motion scenes. Both experiments used the same apparatus, stimuli, and method of paired comparisons that we introduce in the following. The materials and data for all experiments presented in this manuscript are available at https://osf.io/37je6/?view_only=425ecd2c3d3a4edaab89c0d97c7ff2d5

2.1. Method

2.1.1. Participants

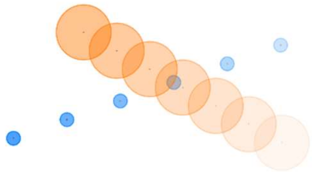
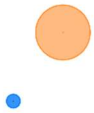
We recruited 26 participants (23 female; age: 18-29 years; mean age: 23.19 years) for Experiment 1a and another 26 participants (21 female; age: 19-35 years; mean age: 24.04 years) for Experiment 1b. All participants were students from the University of Tübingen and participated in exchange for monetary compensation. We ensured that none of the participants participated in multiple of the reported experiments. Participants were treated in accordance with APA standards of ethical treatment and participants provided informed consent prior to the participation.

2.1.2. Apparatus

We ran all experiments on Microsoft Surface Pro tablets with a type cover in order to record participants' responses. Participants sat at an unrestricted viewing distance of about 57 cm to the screen. We tested up to four participants at the same time. Participants were separated by visual shields and could not see each other nor the screen or keyboard of one another.

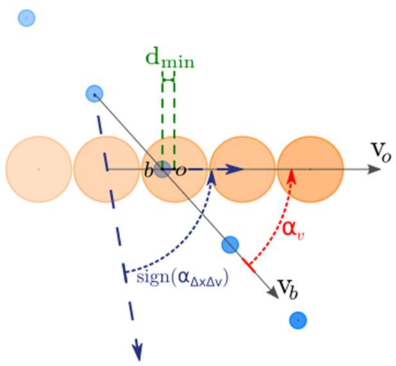
2.1.3. Stimuli

We presented motion scenes that depicted two differently sized and colored discs moving at uniform linear paths for 3 s and passing the moment of minimum distance at the middle of the video clip, thus at 1.5 s. Thus, all motion scenes evolved symmetrically around the point of minimum distance. In all our stimuli, one disc was presented in a blueish color (RGB: 3, 123, 252; 20% transparency; edge: same RGB; 6% transparency) and the other disc was presented in an orangish color (RGB: 255, 119, 0; 50% transparency; edge: same RGB; 25% transparency), see Figure 4. We consider these colors as neutral to interpretation, and they are color-blind distinguishable. Further, the two discs had different sizes: the radius of the orangish disc (r_o) was four times longer than that of the blueish disc (r_b), i.e., $r_o = 4r_b$. Each disc had a black dot in the center to ensure that observers could discern the discs trajectories, und, thus, the observers could unambiguously determine the Motion-OPRA$_1$ categorization. Each video clip was generated and presented at 60 fps for maximum fluidity. All motion scenes began and ended with non-overlapping entities with a minimal separation between disc borders of at least twice the radius of the smaller disc. All throughout this manuscript, we denote the small blue disc by *b* and the large orange disc by *o*.

<table>
<tr><td>
(A) A full stimulus. The 3s film is presented here as seven equally spaced snapshots with temporal fading – the fader the earlier.</td><td>
(B) A frame of the stimulus film in Figure 4A.</td></tr>
<tr><td colspan="2">Figure 4. Appearance of a single stimulus: two moving entities as discs. Entity b is the smaller blue disc; entity o is the larger orange disc. The radii ratio is $r_b\!:\!r_o = 1\!:\!4$</td></tr>
</table>

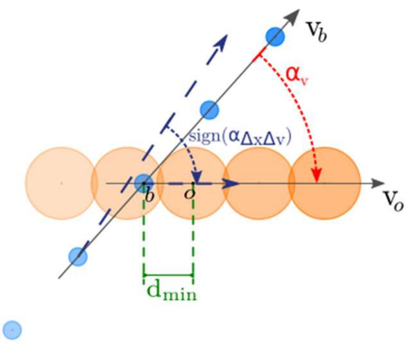
For those motion scenes, the story-based categorization is determined by three variables: d_{min}, α_v, and sgn(α_{DxDv}) – we call them *featural variables* (See Figure 5). The Motion-RCC category of a motion scene is fully determined by the variable d_{min}, which is the distance of the disc centers at their closest approach; the Motion-OPRA$_1$ category of a motion scene is fully determined by α_v and sgn(α_{DxDv}). α_v is the angle between the two trajectories (from v_b to v_o) and sgn(α_{DxDv}) determines how o moves with respect to b. We chose the featural variables by two criteria: first, they must be diagnostic, that is, their values must unequivocally determine each category; second, their value must remain constant for each category (i.e., each story). For that reason, we took the sign and not the angle α_{DxDv}: this angle varies throughout a story, but the sign remains constant. Note that the first two variables d_{min} and α_v are continuous, while the sgn(α_{DxDv}) is discrete.

Scene A



Motion-RCC: mr-co
Motion-OPRA$_1$: mo-bp

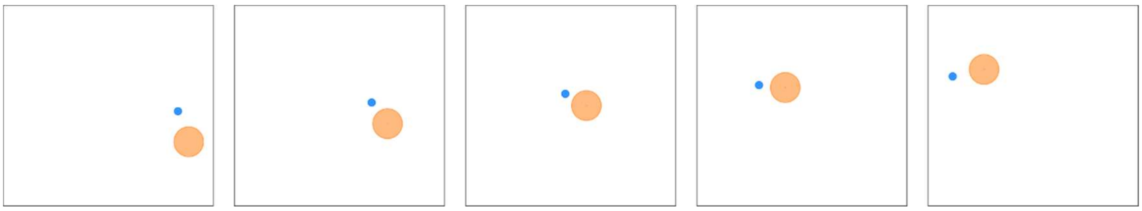
Scene B



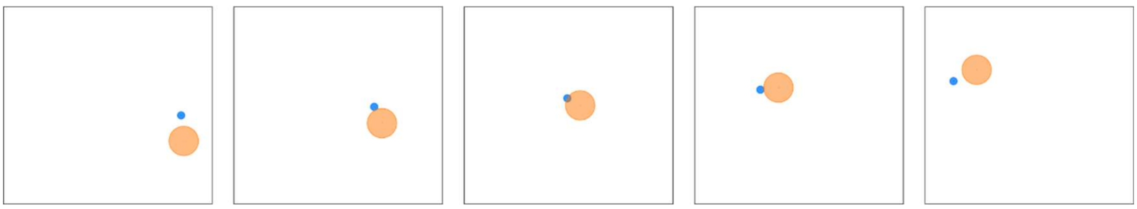
Motion-RCC: mr-dc
Motion-OPRA$_1$: mo-on

Figure 5. Illustration of variables determining the stories of a motion scene:
- d_{min}: the trajectory's minimal distance of the two discs centers, i.e., the distance at the point of maximum approach. If d_{min} is less than $|r_o - r_b|$ the entity b overlaps completely, i.e., the motion scene corresponds to a story mr-co. if d_{min} is greater than $|r_o - r_b|$, but less than $r_o + r_b$ the entity b overlaps partially, i.e., we have a story mr-po. And, if d_{min} greater than $r_o + r_b$ the entities do not overlap, i.e., story mr-dc. d_{min} fully determines the Motion-RCC story.
- α_v: the angle between the velocity vectors of the discs, $\alpha_v = \angle(\overrightarrow{v_b}, \overrightarrow{v_o})$. This angle can be expressed in common language as "o comes from the left" or, equivalently, "b comes from the right" for values in interval (-180°, 0°), and "o comes from the right" or, equivalently, as "b comes from the left" for values in interval (0°, 180°). We avoided values very close to the 0° or 180° boundaries.
- sgn(α_{DxDv}): it is the sign of the angle between the vector of positions difference $\vec{x}_o - \vec{x}_b$ and the vector of velocities difference $\vec{v}_o - \vec{v}_b$. The interpretation of this variable is not as straightforward as α_v. To visually obtain sgn(α_{DxDv}), we must set our reference frame on entity b, and, from that reference frame, we observe the motion of entity o. If, from that viewpoint, o moves counterclockwise around us, then sgn(α_{DxDv}) = +1 (see Scene A). If, otherwise, o moves clockwise around us, then sgn(α_{DxDv}) = -1 (see Scene B). Together with α_v, it determines the Motion-OPRA$_1$ story.

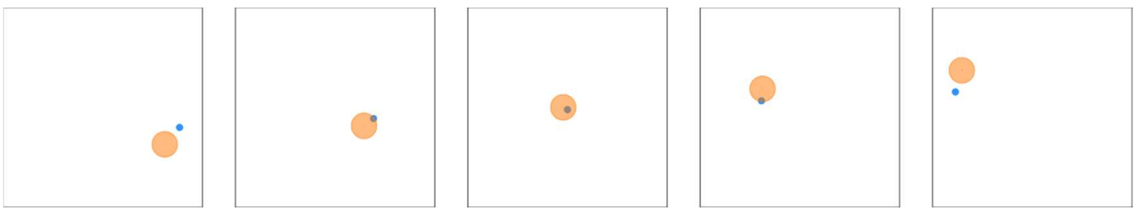
For the present experiments, we chose motion stories that were easily discriminable. We changed, however, the original name of the motion stories, as given in Purcalla Arrufi and Kirsch (2018), into more graspable ones (See Appendix A for a name equivalence table). That is, we named the stories such that all Motion-RCC stories start with the prefix 'mr-' (e.g., mr-po) and all Motion-OPRA$_1$ stories start with the prefix 'mo-' (e.g., mo-bp). Figure 6 depicts examples of the three Motion-RCC stories used in our experiment: mr-no (**n**o **o**verlap), mr-po (**p**artial **o**verlap), and mr-co (**c**omplete **o**verlap). Figure 7 depicts examples of the four Motion-OPRA$_1$ stories used in our experiment: mr-bp, mr-bn, mr-op, and mr-on. The name of the suffix reflects the values of the featural variables α_{v} and sgn(α_{DxDv}). In the first suffix letter, we denote the entity ('o': large orange disc, 'b': small blue disc) that approaches the other from the left – as seen by an external observer. In the second suffix letter, we denote the relative motion of 'o' with respect to 'b': 'p' means that 'b' observes 'o' turning counterclockwise, 'n' means that 'b' observes 'o' turning clockwise.



(a) Story mr-no, 'no overlap' (DC): the entities do not touch, only qualitative relation DC occurs.

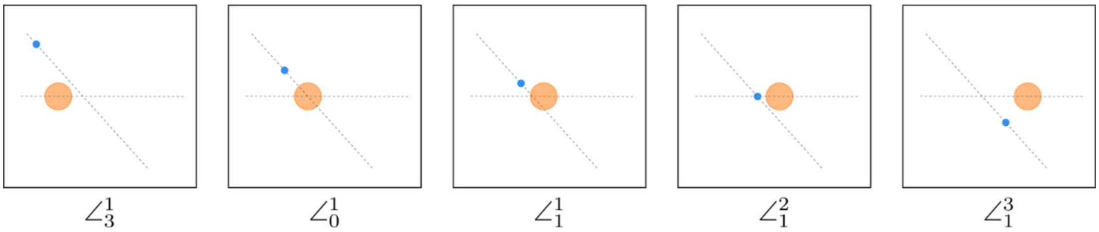


(b) Story mr-po, 'partial overlap' (DC, EC, PO, EC, DC): the entities overlap, relation PO occurs (3rd frame), but there is no complete overlap, i.e., no relation NTPP.

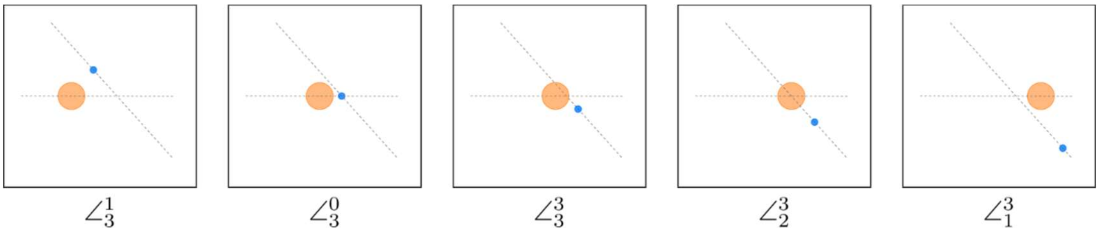


(c) Story mr-co, 'complete overlap' (DC, EC, PO, TPP, NTPP, TPP, PO, EC, DC): the qualitative relation NTPP occurs (3rd frame).

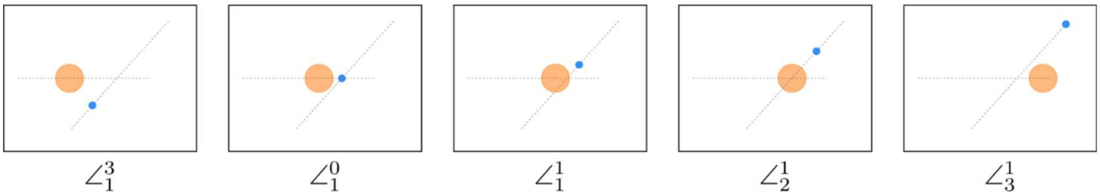
Figure 6. Examples of the Motion-RCC stories used in our experiment. The figure depicts five equally spaced frames from each story.



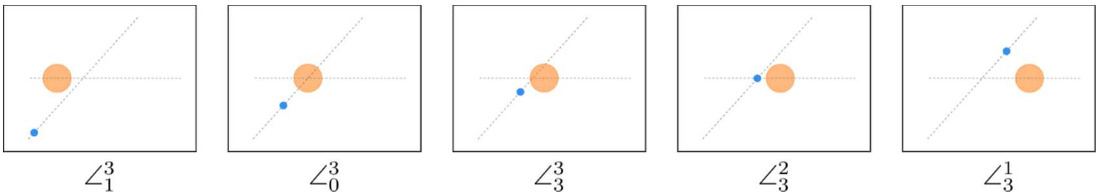
(a) Story mo-bp: b coming from the left of o, o seen turning counterclockwise from b



(b) Story mo-bn: b coming from the left of o, o seen turning clockwise from b



(c) Story mo-op: o coming from the left of b, o seen turning counterclockwise from b



(d) Story mo-on: o coming from the left of b, o seen turning clockwise from b

Figure 7. Examples of the Motion-$OPRA_1$ stories used in our experiment. The figure depicts five equally spaced frames from each story. The trajectories are presented for illustrative purposes only and were not visible in the experiments.

2.1.4. Stimulus Sets

We generated 16 sets of those motion scenes for our experiment (see Appendix B). Each set consisted of one reference stimulus and four transformed stimuli. The reference stimulus was chosen in a way such that applying the same absolute metric change in the two features variables d_{min} and α_v (e.g., d_{min} increased/decreased by 1.25 and α_v increased/decreased by 20°) resulted in four transformed stimuli with one transformed stimulus retaining the Motion-RCC and Motion-$OPRA_1$ categories of the reference stimulus, one transformed stimulus having a different Motion-RCC and Motion-$OPRA_1$ category compared to the reference stimulus, and two transformed stimuli retaining one category and differing in the other category as compared to the reference stimulus (see Figure 8). Whereas d_{min} determined whether the Motion-RCC category changed or not, α_v determined whether the Motion-OPRA$_1$ category changed or not. Note that the third featural variable, *sgn(α_{DxDv})*, was left unchanged for all stimuli.

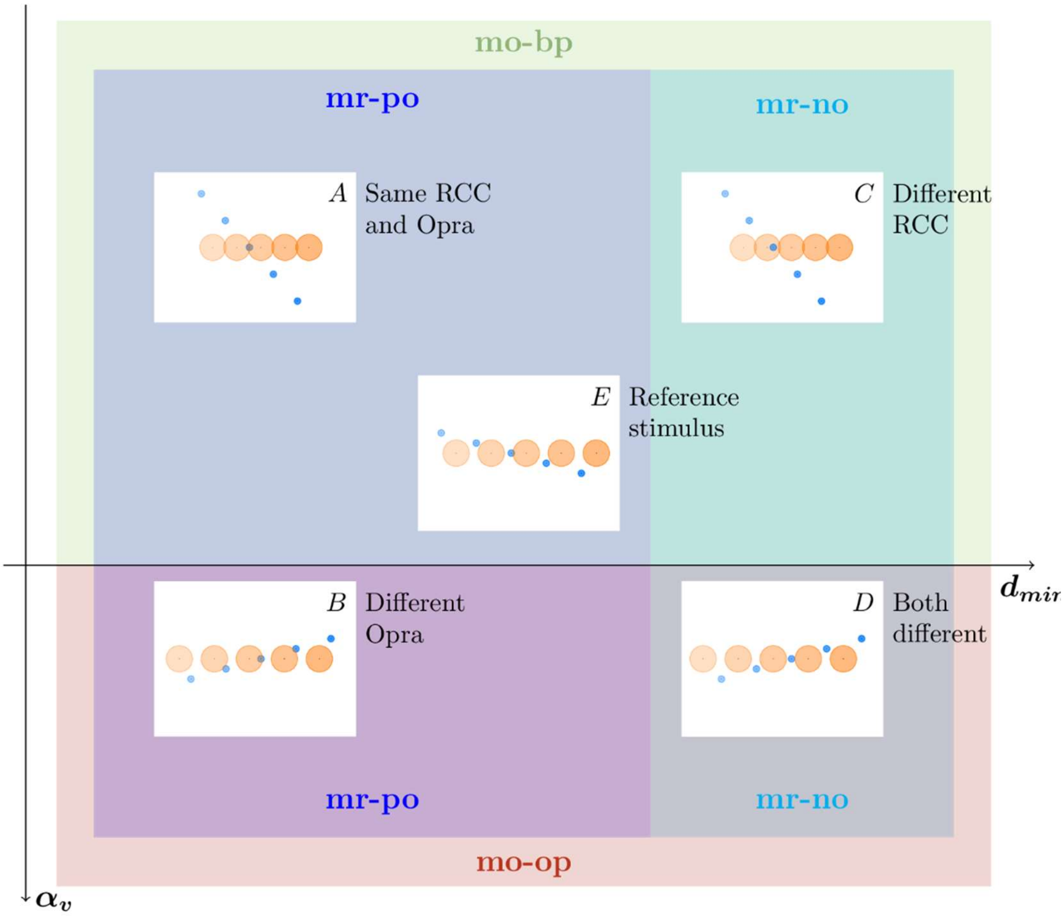


Figure 8. Process of generating a set of motion stimuli: We started by a reference stimulus E and generated the other stimuli from it. We modify the 2 continuous featural variables represented by the axes (α_v is abscissa and d_{min} is ordinate). The variable sgn(α_{DxDv}) is the same for all stimuli in the set.

- stimulus A: the Motion-RCC and Motion-OPRA$_1$ stories are the same as in the reference stimulus E, though, in this stimulus, we modified d_{min} and α_v by the same amount we modified them in the other stimuli.
- stimulus B: Motion-RCC story is the same as in stimuli A and E, we modified d_{min} by the same amount as in stimulus A. Motion-OPRA$_1$ story is different, we modified α_v by the same amount that led from stimulus E to A, but in the other direction.
- stimulus C: Motion-OPRA$_1$ is the same story as in stimulus E, because α_v has the same value as in stimulus A. The Motion-RCC story is different from stimulus E, because d_{min} is increased, but symmetrically with respect to stimulus A. That is, the amount d_{min} is increased between stimulus E and C is the same amount decreased between stimulus E and A.

- stimulus D: It has both different Motion-RCC and Motion-OPRA$_1$ stories from stimulus E. We changed d_{min} as in stimulus C; we changed α_v as in B.

In our experiment, we considered four category border-crossings in Motion-RCC (mr-no ➔ mr-po, mr-po ➔ mr-no, mr-po ➔ mr-co, mr-co ➔ mr-po) and four category border-crossings in Motion-OPRA$_1$ (mo-op ➔ mo-bp, mo-bp ➔ mo-op, mo-on ➔ mo-bn, mo-bn ➔ mo-on) – please note that we varied the category border-crossing in a symmetric manner (e.g., mr-no ➔ mr-po and mr-po ➔ mr-no) to ensure that each category could apply to both the reference stimulus and transformed stimulus to avoid potential biases. Each stimuli set was generated from two category border-crossings, one for Motion-RCC and one for Motion-OPRA$_1$. The combination of border-crossings leads to a total of (4x4) 16 stimulus sets.

To increase the variety in the visual appearance of our stimuli, we applied a random global rotation to each stimuli set (same rotation angle for reference stimulus and four transformed stimuli) for each participant. Thus, each participant saw the stimulus set with a different rotation. Further, we ensured that the disc size was equal across all video clips of the same stimuli set, so that disc size could not be a source of dissimilarity.

2.1.5. Procedure

Participants first provided informed consent and demographic details. Participants then read the instructions about their task, namely, to select which of two motion patterns was more similar to a reference pattern across a large number of trials. In each trial, three empty rectangles with a black outline and a size of 7.6 x 7.6 degree of visual angle appeared after a blank of 1 s. The rectangles were placeholders for the presentation of the motion scenes. The reference pattern was presented in the upper centered rectangle and the comparison patterns in the lower left and right rectangles (see Figure 9). Each video clip lasted for 3 seconds, they were presented synchronously and up to three times successively.

Participants were instructed to press the key “f” if the lower left motion pattern was more similar to the reference or to press the key “j” if the lower right motion pattern was more similar to the reference. They were also instructed that they would see the motion patterns up to three times such that they would have enough time to reach a decision, but that they should answer as soon as they decided and that they need not wait until the patterns played for three times. Following each trial, we presented a fixation cross and participants could take a self-paced break and continue to the next trial by pressing the spacebar. After every 10 percent progress of trials, we also informed participants about their total progress above the fixation cross within this display. Following all trials, we thanked the participants for their participation and they received monetary compensation for their participation.

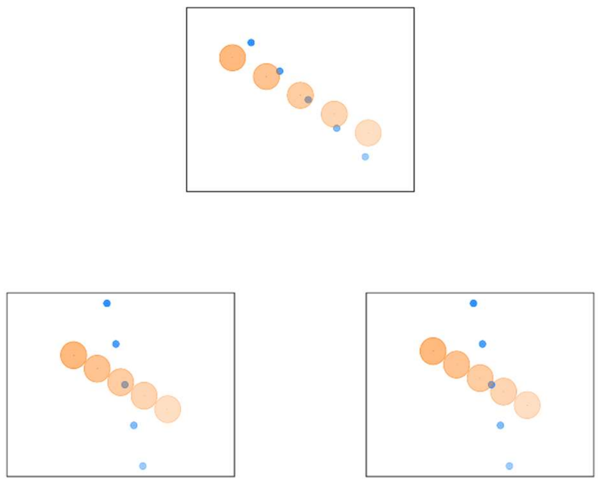


Figure 9. Arrangement of stimuli on the tablet screen in the perception experiments. The key is that the reference (above) and comparison (below) stimuli were presented simultaneously.

Experiments 1a and 1b differed regarding which motion scenes were presented for comparison with the reference stimulus in the lower two panels. In Experiment 1a, the set of comparison stimuli for each reference stimulus comprised the four transformed stimuli. That is, all comparison stimuli had the same absolute metric changes applied but differed with respect to whether the Motion-RCC and/or Motion-OPRA$_1$ category was retained or changed. Thus, participants performed 6 different pairwise comparisons, all possible pairs of 4 transformed stimuli $\binom{4}{2}$ for each stimuli set. In Experiment 1b, we added a copy of the reference stimulus to the set of comparison stimuli. Thus, participants performed 10 different pairwise comparisons, all possible pairs of 4 transformed stimuli + 1 reference stimulus, $\binom{5}{2}$ for each stimuli set. In each of the two experiments, participants started with a full set of pairwise comparisons with a practice reference stimulus, that is there were 6 practice trials in Experiment 1a and 10 practice trials in Experiment 1b. Thereafter, participants performed the experimental trials. In Experiment 1a, we presented the 16 stimuli sets twice (with a new random global rotation angle applied) leading to 16 (stimuli sets) x 6 (pairwise comparisons per stimuli set) x 2 (repetitions) = 192 experimental trials. In Experiment 1b, we presented each stimuli set once resulting in 16 (stimuli sets) x 10 (pairwise comparisons per stimuli set) = 160 experimental trials. In both experiments, the experimental trials were presented randomly intermixed with the restriction that there was at least one intervening trial before the repetition of the same stimulus set. For each trial, we randomly determined which of the two comparison stimuli was presented in the lower left or lower right panel.

2.1.6. Determining Similarity: Paired Comparisons

With the present experiments, we aimed at investigating whether the properties underlying the categorizations according to Motion-RCC and Motion-$OPRA_1$ might also influence the human representation of motion scenes for either perception or memorization. Therefore, we measured how similar participants perceived the transformed stimuli to the respective reference stimulus. If the categorization according to Motion-RCC and Motion-$OPRA_1$ is irrelevant to human cognition, all transformed stimuli should be perceived as equally similar to the reference stimulus, because we applied the same absolute metric changes to all of them. If such categorizations do influence human cognition, however, we should see reduced similarity scores for those transformed stimuli where the Motion-RCC and/or Motion-$OPRA_1$ category was changed.

We took care not to bias our participants towards specific aspects of similarity; for that reason, we applied the method of paired comparisons. For each trial, participants saw one reference stimulus and two comparison stimuli; then, they had to choose which of the two comparison stimuli they perceived as being more similar to the reference stimulus. From those choices, we obtained the ratio scale “similarity to reference stimulus” by fitting the Bradley-Terry-Luce (BTL) model with the R package eba (Wickelmaier, 2020; Wickelmaier & Schmid, 2004). That is, using the fitted ratio scale, we could both measure whether our comparison stimuli were judged as differing in their perceived similarity to the reference stimulus, and, if so, to what extent.

2.2. Experiment 1a: Results and Discussion

From the responses of the pair comparisons, we obtained the ratio scale “similarity to reference stimulus” by fitting the Bradley-Terry-Luce (BTL) model with the R package eba (Wickelmaier, 2020). To increase the comparability of the results across the presented experiments, we normalized the resulting ratio scale such that the stimulus with retained Motion-OPRA$_1$ and Motion-RCC categories took the value 1 (see Figure 10) in all experiments. The BTL model provided an acceptable fit to the observed data as indicated by a non-significant goodness of fit test, $\chi^2(3) = 7.68$, $p = .053$. There was a significant effect of motion category, $\chi^2(3) = 111.69$, $p < .001$, indicating that motion category affected participants’ similarity judgements. The similarity scale values including the 95% confidence intervals are depicted in Figure 10.

There were two main findings. First, the categories defined in Motion-OPRA$_1$ and Motion-RCC had an additive effect in participants’ similarity judgements: Even though the amount of the metric manipulation of the underlying motion properties was comparable for all comparison stimuli, a change to either the Motion-OPRA$_1$ or Motion-RCC category resulted in decreased similarity judgments. When both categories changed together, we observed an additive effect resulting in the strongest decrease in perceived similarity to the reference stimulus. Second, as evident from Figure 10, a change in the Motion-OPRA$_1$ category had a stronger influence on the similarity judgments than a change in the Motion-RCC category.

One potential limitation of Experiment 1a was that the set of comparison stimuli comprised only transformed stimuli, i.e., stimuli with metric changes, potentially leading to an overestimation of the role of motion stories on visual perception as compared to situations that also contain stimuli that are more like the reference stimulus. Therefore, we performed Experiment 1b, where we added a copy of the reference stimulus to the set of comparison stimuli.

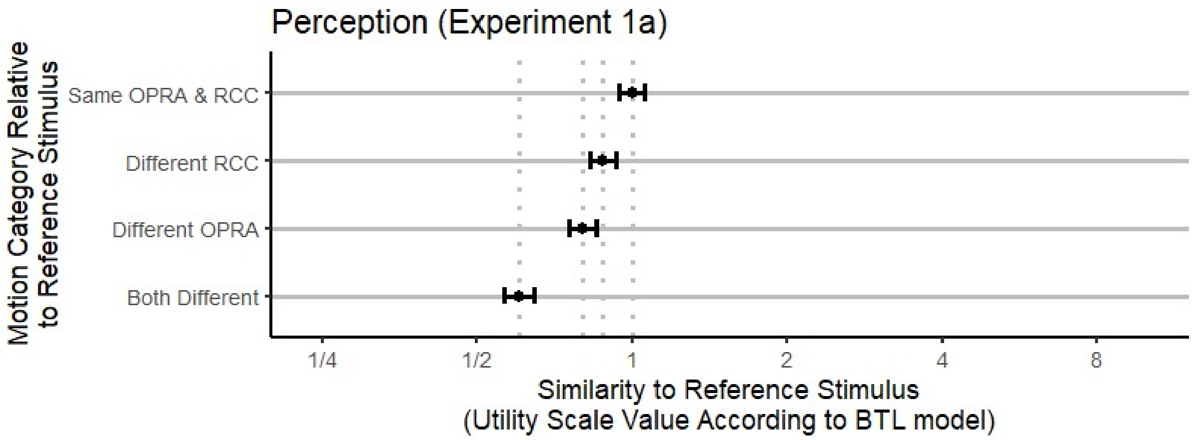


Figure 10. BTL similarity scale values and their 95% confidence intervals derived from the paired comparisons of Experiment 1a.

2.3. Experiment 1b: Results and Discussion

We again derived the ratio scale “similarity to reference stimulus” using the BTL model which provided a good fit to the observed data as indicated by a non-significant goodness of fit test, $\chi^2(6) = 5.44$, $p = .489$. There was again a significant effect of motion category, $\chi^2(4) = 1664.58$, $p < .001$, indicating that motion category affected participants’ similarity judgements. The similarity scale values including the 95% confidence intervals are depicted in Figure 11.

Replicating Experiment 1a, we observed that participants used the categories provided by the Motion-OPRA$_1$ and Motion-RCC categorizations for their similarity judgments, even though the set of comparison stimuli now also included a copy of the reference stimulus. Motion-OPRA$_1$ again had a larger impact on the similarity judgments than Motion-RCC with both categories affecting the similarity judgements in an additive manner.

The present experiment also demonstrated that participants judged the stimulus that was identical to the reference stimulus to be more than eight times more similar to the reference stimulus than the transformed stimulus that had the same Motion-OPRA$_1$ and

Motion-RCC categories as the reference stimulus. This shows that participants relied not only on the categories defined by Motion-OPRA$_1$ and Motion-RCC for their similarity judgements, but that they also used additional stimulus features that are not covered by the two categorizations.

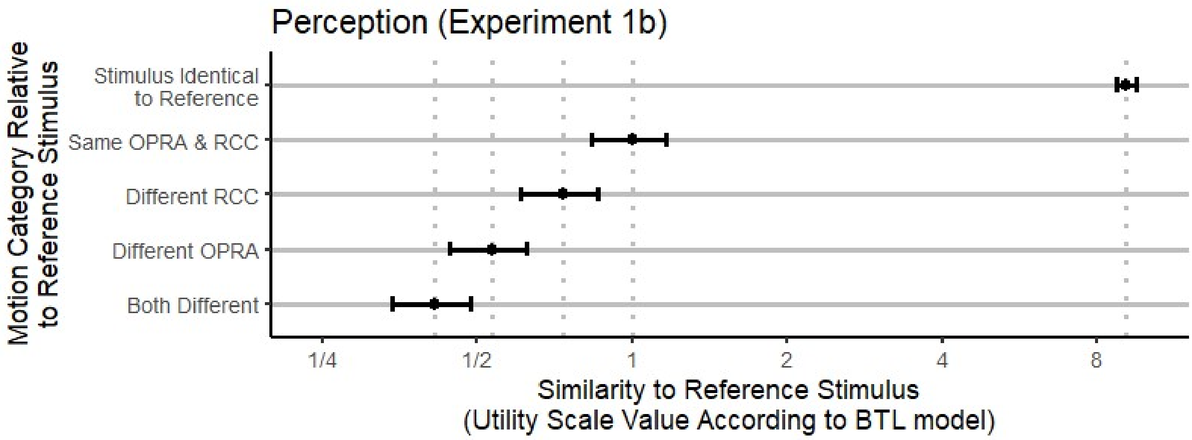


Figure 11. BTL similarity scale values and their 95% confidence intervals derived from the paired comparisons of Experiment 1b.

3. Experimental Set 2: Memory

Experimental set 1 demonstrated that the properties underlying the Motion-RCC and Motion-OPRA$_1$ categories influence human perception. Therefore, in a next step, we investigated whether this accounts also for human short-term memory. We considered this question particularly interesting because human short-term memory operates under strict capacity limitation for static scenes (e.g., Cowan, 2001; Luck & Vogel, 1997). Thus, we deemed it interesting to determine whether the properties underlying Moton-RCC and/or Motion-OPRA$_1$ are also represented within the limited short-term memory, or whether those properties might be partially or totally dropped from perception to memory, rendering some or all of those categories irrelevant for human short-term memory. Therefore, we performed two additional experiments that were similar to experimental series 1, but they slightly differed in procedure, in order to investigate short-term memory rather than perception.

3.1. Method

3.1.1. Participants

We recruited 26 participants (19 female; age: 18-30 years; mean age: 24.08 years; one participant did not provide descriptive details) for Experiment 2a and another 26 participants (21 female; age: 19-31 years; mean age: 24.12 years) for Experiment 2b. All participants were students from the University of Tübingen and participated in exchange for monetary compensation. We ensured that none of the participants participated in multiple of the reported experiments. Participants were treated in accordance with APA standards of ethical treatment and participants provided informed consent prior to the participation.

3.1.2. Apparatus and Stimuli

We used the same apparatus and stimuli as in experimental set 1.

3.1.3. Procedure

We used the same procedure as in experimental series 1 except for the following modifications. Within each trial, there was a blank of 1 s, followed by the reference stimulus being presented for 3 s in the center of the screen. Thereafter, there was another blank of 1 s ensuring that participants had to respond based on their memory rather than perception. Following this maintenance phase, the comparison stimuli appeared simultaneously on the left and right side of the screen (see Figure 12). They were presented twice for 3 s each, thus for a maximum of 6 s. Participants responded with the keys “f” and “j” as in experimental series 1. They were instructed that they would see the comparison patterns up to two times such that they would have enough time to reach a decision, but that they should answer as soon as they decided and that they need not to wait until the patterns played for two times. The motion scenes were again presented within empty placeholder rectangles subtending 7.6 x 7.6 degree of visual angle.

Just as in experimental set 1, Experiments 2a and 2b differed regarding the motion scenes that were presented for comparison with the reference stimulus. In Experiment 2a, it comprised the four transformed stimuli and in Experiment 2b it comprised the four transformed stimuli and a copy of the reference stimulus. The number of practice trials and experimental trials in Experiments 2a and 2b was identical to Experiments 1a and 1b, respectively.

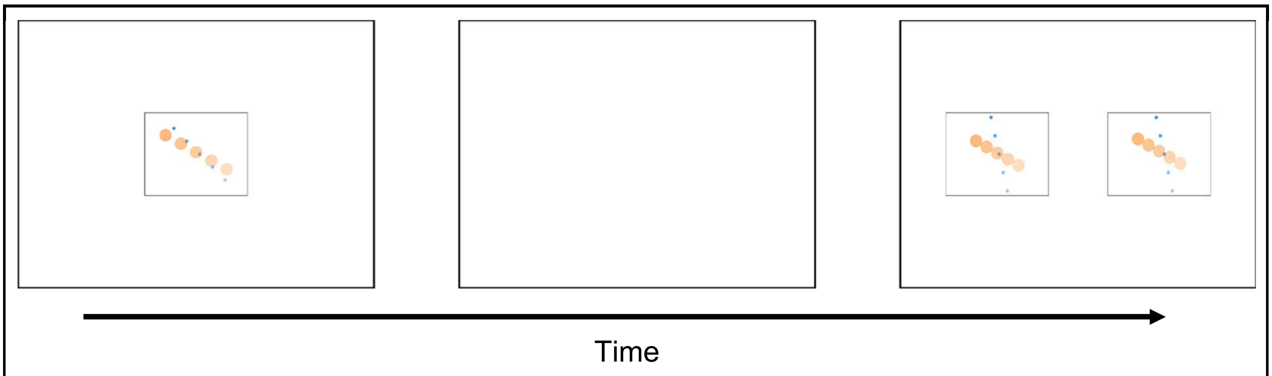


Figure 12. Arrangement of stimuli on the tablet screen in the memory experiments. The key is that the reference and compared stimuli were presented sequentially in time with a blank pause in between.

3.2. Experiment 2a: Results and Discussion

As in Experiment 1a, we derived the ratio scale “similarity to reference stimulus” using the BTL model which provided a good fit to the observed data as indicated by a non-significant goodness of fit test, $\chi^2(3) = 5.43$, $p = .143$. There was again a significant effect of motion category, $\chi^2(3) = 780.17$, $p < .001$, indicating that motion category affected participants’ similarity judgements. The similarity scale values including the 95% confidence intervals are depicted in Figure 13.

Similar to our experiment on perception (Experiment 1a), participants were sensitive to changes in Motion-OPRA$_1$ within the present short-term memory experiment. That is, a change in the Motion-OPRA$_1$ category led to a reduced similarity judgement. In contrast to perception, however, the effect of Motion-RCC on participants’ judgements was largely reduced. For stimuli differing in their Motion-OPRA$_1$ category from the reference stimulus, Motion-RCC had no additional effect on the similarity judgements. Only for those stimuli with retained Motion-OPRA$_1$ category, changing the Motion-RCC category caused a small but significant decrease in similarity judgments.

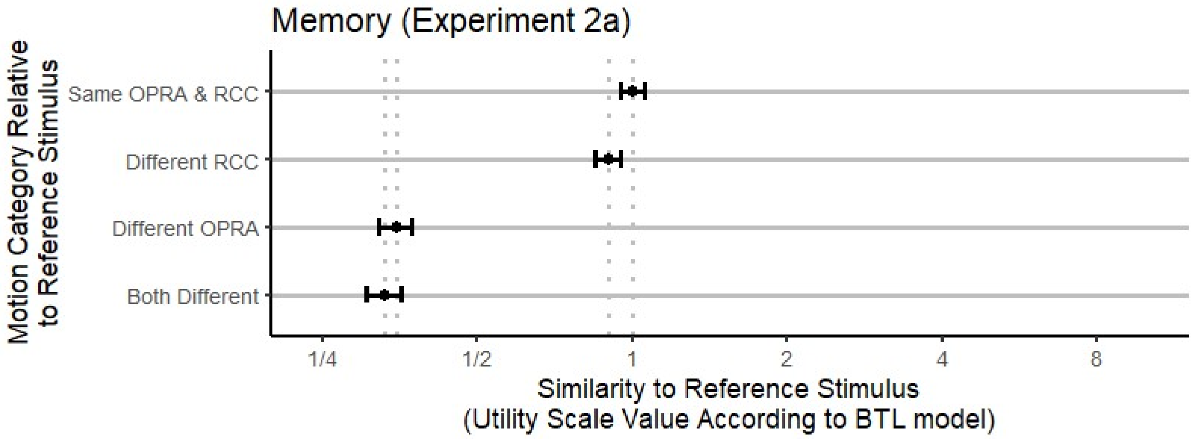


Figure 13. BTL similarity scale values and their 95% confidence intervals derived from the paired comparisons of Experiment 2a.

3.3. Experiment 2b: Results and Discussion

We again derived the ratio scale “similarity to reference stimulus” using the BTL model which provided a good fit to the observed data as indicated by a non-significant goodness of fit test, $\chi^2(6) = 6.79$, $p = .341$. There was again a significant effect of motion category, $\chi^2(4) = 1515.97$, $p < .001$, indicating that motion category affected participants’ similarity judgements. The similarity scale values including the 95% confidence intervals are depicted in Figure 14.

We observed an influence of Motion-OPRA$_1$ on participants’ similarity judgements. In contrast to Experiment 2a, however, we observed no effect of Motion-RCC on the similarity judgements. That is, whereas Experiment 2a demonstrated a small influence of Motion-RCC, at least when the Motion-OPRA$_1$ category was retained, this influence disappeared completely once we added the reference stimulus to the set of comparison stimuli in Experiment 2b. Thus, the influence of Motion-RCC on human short-term memory might either be only observed under very specific conditions, or it might not be reliable after all.

Just like in perception (Experiment 1b), the similarity judgment for the stimulus that was identical to the reference stimulus showed the highest similarity judgements. Thus, also when using their short-term memory, the participants relied on additional stimulus features that are not covered by the Motion-OPRA$_1$ categorization for their similarity judgements.

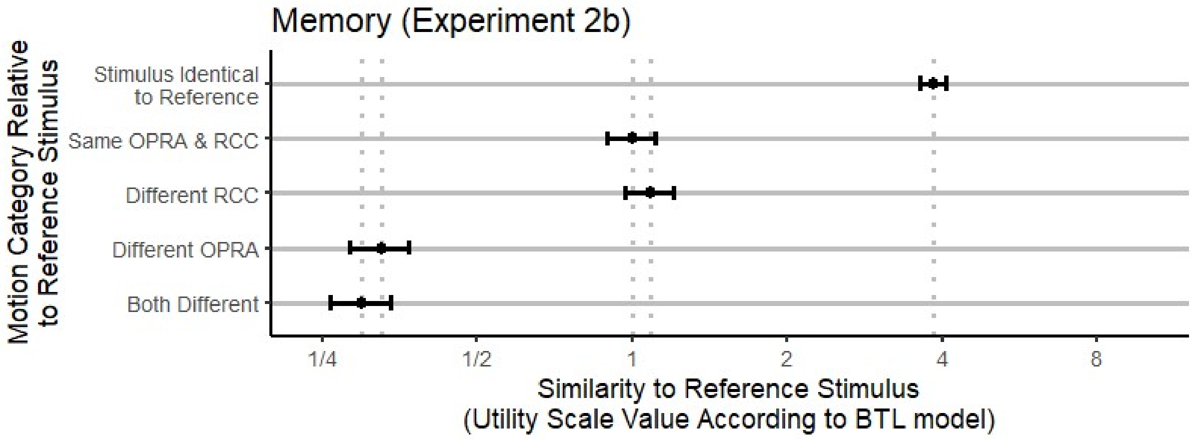


Figure 14. BTL similarity scale values and their 95% confidence intervals derived from the paired comparisons of Experiment 2b.

4. General Discussion

We asked whether motion categorizations designed in artificial intelligence can inform the psychological understanding of human perception and memory of motion scenes. Thus, we constructed simple motion scenes based on the two story-based categorizations Motion-RCC and Motion-OPRA$_1$. These reference scenes depicted two objects moving on linear trajectories with a constant speed. We then transformed each reference scene into four transformed scenes reflecting changes of no, a single, or both Motion-RCC/Motion-OPRA$_1$ categories. Notably, those transformations consisted of metric changes of the same absolute amount. In four experiments, we asked human participants to judge the subjective similarity of those transformed scenes to the reference scenes by using paired comparisons. Suppose changes in the Motion-RCC or Motion-OPRA$_1$ categories are irrelevant to human cognition. In that case, all transformed scenes should result in the same subjective similarity judgments due to the same absolute metric change being applied. If the Motion-RCC or Motion-OPRA$_1$ categorizations are relevant to human cognition, however, changes in those categories should influence similarity judgments.

We found that both story-based categorizations (Motion-RCC and Motion-OPRA$_1$) affected the similarity judgments when both the reference stimulus and the transformed stimulus were visible simultaneously, thus during human perception (Experiments 1a and 1b). Changes in both Motion-RCC and Motion-OPRA$_1$ categories resulted in reduced similarity judgments. Thus, the regional relations (i.e., extent of overlap; Motion-RCC) and the directional relations (Motion-OPRA$_1$) between the moving objects seem to influence participants' perception.

Interestingly, we observed a different pattern of results when participants based their similarity judgments on short-term memory (Experiments 2a and 2b). When participants memorized the reference stimulus first and then compared the transformed stimulus to their memory representation, there was a marked influence of Motion-OPRA$_1$ only. Thus, participants based their similarity judgments on the directional relations between the moving objects (Motion-OPRA$_1$). Regarding regional relations (Motion-RCC), we observed only a little impact on similarity judgments in Experiment 2a when the Motion-OPRA$_1$ category was unchanged, and thus Motion-RCC was the only distinguishing feature. However, we did not replicate this effect of Motion-RCC in Experiment 2b, and Motion-RCC had no additional influence when Motion-OPRA$_1$ changed, indicating that Motion-RCC was rather irrelevant for memory-based similarity judgments.

For the first time, we show that the Motion-RCC and Motion-OPRA$_1$ categorizations developed in the context of artificial intelligence and qualitative reasoning (Dylla et al., 2017; Gantner et al., 2008; Lücke et al., 2011; Purcalla Arrufi & Kirsch, 2018) are cognitively plausible. That is, the properties underlying those qualitative representations did impact human cognition. At the same time, our study shows that one needs to be cautious when using terms such as *cognitively plausible*. Imagine we had performed the perceptual experiments (Experiments 1a and 1b) only. Based on the finding that Motion-RCC categorizations influenced the similarity judgments, one could conclude that the human representation of motion scenes includes regional representation similar to the overlaps defined in Motion-RCC. Because we also performed a short-term memory task (Experiments 2a and 2b), we know this does not seem to be the case. Whereas regional relations (Motion-RCC) might be represented within a short-lived high-capacity sensory iconic representation (Dick, 1974; Phillips, 1974), this might not be the case for short-term memory, which is prone to much tighter capacity limitations (Phillips, 1974). Our results indicate that participants dropped the regional relations defined by Motion-RCC when storing the motion scenes within short-term memory. Therefore, our results suggest that calling specific categorizations cognitively plausible might be misleading. It might be better to relate to the respective cognitive process directly, such as calling Motion-RCC perceptually plausible and Motion-$OPRA_1$ plausible concerning human perception and short-term memory.

Our results also have direct implications for psychology. First, our findings support the idea that human perception and memory represent visual scenes based not only on absolute metric values but also on spatial relations and categorizations (Huttenlocher et al., 1991, 2000; Jiang et al., 2000; Papenmeier & Huff, 2014; Sampaio & Wang, 2010, 2017). Second, our results indicate that whereas spatial relations might be relevant for both human perception and human short-term memory, the actual spatial representations underlying both systems might differ. Regarding this difference, there are at least two plausible alternatives. Some information (such as the regional representations similar to Motion-RCC) might be dropped when transferring the content from perception to short-term memory, or both systems might be based on entirely different spatial representations. Based on our study, we cannot decide between the two alternatives. However, we speculate that the former might be true because some perceptual spatial relations can directly influence spatial representations in short-term memory even if they are task-irrelevant. For example, Jiang et al. (2004) demonstrated that a task-irrelevant perceptual grouping of objects as rows or columns can directly affect the spatial representation of objects in short-term memory.

The present study investigated whether the story-based categorizations Motion-RCC and Motion-$OPRA_1$ are relevant to human perception or human short-term memory. As we have shown, this is the case for human perception and short-term memory in the case of Motion-$OPRA_1$ and for human perception in the case of Motion-RCC. Our findings thus open up two important new research questions for future research. First, do the regional and directional relations described by Motion-RCC and Motion-$OPRA_1$ directly affect human spatial representations? As an alternative, other spatial features, such as which object crosses first, might guide human spatial representations of motion scenes and might only correlate with the Motion-RCC or Motion-OPRA$_1$ categorizations studied in our present experiments. Second, do participants represent the motion scenes as stories or static snapshots? We studied the influence of the story-based categorizations Motion-RCC and Motion-OPRA$_1$ on the perception and memory of motion scenes. Those story-based categorizations define stories as transitions across multiple states of regional or directional relations. We found evidence for the influence of those story-based categorizations on human perception and short-term memory. However, our experiments were not designed to distinguish between participants representing the motion scenes as complete stories or single static snapshots. We speculate that participants might form story-based representations of the motion scenes based on previous research. First, humans can represent dynamic properties of motion scenes, such as dynamic spatial configurations (Papenmeier et al., 2012; Sun et al., 2015), in short-term memory. Second, theories on event cognition propose that humans represent their dynamic environment in the form of events (e.g., Loschky et al., 2019; Radvansky & Zacks, 2014; Zacks, 2020; Zacks et al., 2007), which might have properties similar to the proposed story-based spatial representations. Future research needs to investigate whether participants indeed form story-based spatial representations and to what extent the transitions between different RCC or OPRA$_1$ states are reflected within such representations.

Collision avoidance is an essential daily skill for humans (e.g., Huber et al., 2014); and as such, it is also a coveted skill for human-robot interaction (e.g., Shiomi et al., 2014), specifically for driving assistant systems and, ultimately, for autonomous driving. Given that Motion-RCC categorizes the degree of overlap between entities, that is, the degree of collision of both entities, it might be unexpected that Motion-RCC had a lower effect than Motion-OPRA1 on human motion categorization. Since our results provide the first evidence for the lower importance of Motion-RCC than Motion-OPRA1, it is up to future work to determine whether our categorization findings generalize to other motion stimuli.

While our research showed that some spatial representations from artificial intelligence seem to have some connection to the mental mechanisms of humans, it is important to note that they are far from representing the whole range of human understanding of motion. We also observed that participants judged the identical stimulus as more similar to the reference stimulus than the stimulus with retained Motion-RCC and Motion-OPRA$_1$ properties. Thus, our participants used information beyond the properties underlying Motion-RCC and Motion-OPRA$_1$ to represent the motion scenes. Besides additional categorical information, it seems plausible that some metric information also plays a role in representing motion scenes. Further knowledge about that information would provide an exciting challenge for AI knowledge representation mechanisms that are usually only symbolic or only numeric. A tight integration of metric and categorical information could open many new possibilities for representing and reasoning about real-world situations. We encourage further interdisciplinary research to investigate these mechanisms to develop well-defined models of spatial representations. Such progress would also be important for use cases such as human-aware robot navigation. We can hardly expect a robot to behave socially if it has a totally different representation of what is going on than the humans around it.

Given that many artificial intelligence applications should eventually be applied in real life, it would be interesting to see whether our findings generalize to scenarios with a higher ecological validity or even real-world environments. Our present stimuli (two colored discs) differ from real-world entities such as pedestrians and cars not only in appearance, but also in the variability of motion and trajectory properties and the top-view presentation of flat 2D objects compared to the first-person view of voluminous 3D objects. In our present research, we choose to study categorizations with entities free from semantic influence and from a perspective that provides clear indications of spatial overlap and directional relations. When considering Motion-RCC, increasing ecological validity poses some challenges that need to be solved. For example, overlapping objects cannot occur with rigid objects because these would imply collisions of the moving entities. One solution could be to study Motion-RCC with respect to the two-dimensional depiction of the three-dimensional environment, such as if viewed through a camera. One could argue that overlaps could occur within such a depiction, such as when watching a car moving behind a truck.

To summarize, we studied the influence of motion categorizations as derived from qualitative representation in artificial intelligence on human perception and memory. We controlled for absolute metric changes and found marked influences of those categorizations on similarity judgments of motion scenes. Thus, human perception and short-term memory utilize some of the properties underlying those qualitative representations. Given this link between human cognition and qualitative reasoning, we argue for a continued and close multidisciplinary approach to investigating the spatial representation of motion scenes. We are confident that this will help inform both psychological theories of human cognition and artificial intelligence models, providing new grounds that help move both fields forward.

Author's contributions

All authors developed the study concept and contributed to the study design. JPA developed the stimulus materials and FP programmed the experimental environment. FP was responsible for data collection and the analysis of the experiments. All authors were responsible for the interpretation of the results and drafted the manuscript. All authors approved the final version of the manuscript for submission.

Acknowledgments

We thank Lisa Arndt and Nele Vagt for their help with data collection. Juan Purcalla Arrufi thanks heartfully his wife for her economic, emotional, and prayer support throughout this research.

References

Banich, M. T., & Federmeier, K. D. (1999). Categorical and metric spatial processes distinguished by task demands and practice. *Journal of Cognitive Neuroscience*, *11*(2), 153–166. https://doi.org/10.1162/089892999563292

Basili, P., Sağlam, M., Kruse, T., Huber, M., Kirsch, A., & Glasauer, S. (2013). Strategies of locomotor collision avoidance. *Gait & Posture*, *37*(3), 385–390. https://doi.org/10.1016/j.gaitpost.2012.08.003

Cohen, H., & Lefebvre, C. (2005). Bridging the category divide. In H. Cohen & C. Lefebvre (Eds.), *Handbook of Categorization in Cognitive Science* (pp. 1–15). Elsevier.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.

Dick, A. O. (1974). Iconic memory and its relation to perceptual processing and other memory mechanisms. *Perception & Psychophysics*, *16*(3), 575–596. https://doi.org/10.3758/BF03198590

Dylla, F., Lee, J. H., Mossakowski, T., Schneider, T., Delden, A. V., Ven, J. V. D., & Wolter, D. (2017). A survey of qualitative spatial and temporal calculi: Algebraic and computational properties. *ACM Computing Surveys*, *50*(1), 7:1-7:39. https://doi.org/10.1145/3038927

Fehd, H. M., & Seiffert, A. E. (2008). Eye movements during multiple object tracking: Where do participants look? *Cognition*, *108*, 201–209. https://doi.org/10.1016/j.cognition.2007.11.008

Fehd, H. M., & Seiffert, A. E. (2010). Looking at the center of the targets helps multiple object tracking. *Journal of Vision*, *10*(4:19), 1–13. https://doi.org/10.1167/10.4.19

Gantner, Z., Westphal, M., & Wölfl, S. (2008). GQR - A fast reasoner for binary qualitative constraint calculi. *AAAI Workshop on Spatial and Temporal Reasoning*.

Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 31–47. https://doi.org/10.1037/0096-1523.33.1.31

Huber, M., Su, Y.-H., Krüger, M., Faschian, K., Glasauer, S., & Hermsdörfer, J. (2014). Adjustments of speed and path when avoiding collisions with another pedestrian. *PLOS ONE*, *9*(2), e89589. https://doi.org/10.1371/journal.pone.0089589

Huff, M., Papenmeier, F., Jahn, G., & Hesse, F. W. (2010). Eye movements across viewpoint changes in multiple object tracking. *Visual Cognition*, *18*, 1368–1391. https://doi.org/10.1080/13506285.2010.495878

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, *98*(3), 352–376. https://doi.org/10.1037/0033-295X.98.3.352

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, *129*, 220–241. https://doi.org/10.1037/0096-3445.129.2.220

Jiang, Y., Chun, M. M., & Olson, I. R. (2004). Perceptual grouping in change detection. *Perception & Psychophysics*, *66*, 446–453.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 683–702. https://doi.org/10.1037//0278-7393.26.3.683

Klinghammer, M., Blohm, G., & Fiehler, K. (2015). Contextual factors determine the use of allocentric information for reaching in a naturalistic scene. *Journal of Vision*, *15*(13:24), 1–13. https://doi.org/10.1167/15.13.24

Klinghammer, M., Blohm, G., & Fiehler, K. (2017). Scene configuration and object reliability affect the use of allocentric information for memory-guided reaching. *Frontiers in Neuroscience*, *11:204*. https://doi.org/10.3389/fnins.2017.00204

Klinghammer, M., Schütz, I., Blohm, G., & Fiehler, K. (2016). Allocentric information is used for memory-guided reaching in depth: A virtual reality study. *Vision Research*, *129*, 13–24. https://doi.org/10.1016/j.visres.2016.10.004

Kosslyn, S. M. (1987). Seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review*, *94*(2), 148–175. https://doi.org/10.1037/0033-295X.94.2.148

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. The University of Chicago Press.

Loschky, L. C., Larson, A. M., Smith, T. J., & Magliano, J. P. (2019). The Scene Perception & Event Comprehension Theory (SPECT) Applied to Visual Narratives. *Topics in Cognitive Science*, tops.12455. https://doi.org/10.1111/tops.12455

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Lücke, D., Mossakowski, T., & Moratz, R. (2011). Streets to the OPRA-Finding your destination with imprecise knowledge. *IJCAI Workshop on Benchmarks and Applications of Spatial Reasoning*, 25–32.

Medin, D. L., & Heit, E. (1999). Chapter 3—Categorization. In B. M. Bly & D. E. Rumelhart (Eds.), *Cognitive Science* (pp. 99–143). Academic Press. https://doi.org/10.1016/B978-012601730-4/50005-6

Meyerhoff, H. S., Papenmeier, F., & Huff, M. (2017). Studying visual attention using the multiple object tracking paradigm: A tutorial review. *Attention, Perception, & Psychophysics*, *79*, 1255–1274. https://doi.org/10.3758/s13414-017-1338-1

Mossakowski, T., & Moratz, R. (2012). Qualitative reasoning about relative direction of oriented points. *Artificial Intelligence*, *180–181*, 34–45. https://doi.org/10.1016/j.artint.2011.10.003

Murphy, G. L. (2002). *The big book of concepts*. MIT Press.

Papenmeier, F., & Huff, M. (2014). Viewpoint-dependent representation of contextual information in visual working memory. *Attention, Perception, & Psychophysics*, *76*, 663–668. https://doi.org/10.3758/s13414-014-0632-4

Papenmeier, F., Huff, M., & Schwan, S. (2012). Representation of dynamic spatial configurations in visual short-term memory. *Attention, Perception, & Psychophysics*, *74*, 397–415. https://doi.org/10.3758/s13414-011-0242-3

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*, 283–290.

Purcalla Arrufi, J., & Kirsch, A. (2018). Motion categorisation: Representing velocity qualitatively. *Cognitive Systems Research*, *52*, 117–131. https://doi.org/10.1016/j.cogsys.2018.06.005

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179–197. https://doi.org/10.1163/156856888X00122

Radvansky, G. A., & Zacks, J. M. (2014). Event model structure and processing. In *Event Cognition*. Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199898138.001.0001/acprof-9780199898138

Randell, D. A., Cui, Z., & Cohn, A. G. (1992). A spatial logic based on regions and connection. *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning*, 165–176.

Renz, J., Rauh, R., & Knauff, M. (2000). Towards cognitive adequacy of topological spatial relations. In C. Freksa, C. Habel, W. Brauer, & K. F. Wender (Eds.), *Spatial Cognition II* (pp. 184–197). Springer. https://doi.org/10.1007/3-540-45460-8_14

Sampaio, C., & Wang, R. F. (2010). Overcoming default categorical bias in spatial memory. *Memory & Cognition*, *38*(8), 1041–1048. https://doi.org/10.3758/MC.38.8.1041

Sampaio, C., & Wang, R. F. (2017). The cause of category-based distortions in spatial memory: A distribution analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 1988–1992. https://doi.org/10.1037/xlm0000424

Shiomi, M., Zanlungo, F., Hayashi, K., & Kanda, T. (2014). Towards a socially acceptable collision avoidance for a mobile robot navigating among pedestrians using a pedestrian model. *International Journal of Social Robotics*, *6*(3), 443–455. https://doi.org/10.1007/s12369-014-0238-y

Sun, Z., Huang, Y., Yu, W., Zhang, M., Shui, R., & Gao, T. (2015). How to break the configuration of moving objects? Geometric invariance in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 1247–1259. https://doi.org/10.1037/xhp0000086

Timm, J. D., & Papenmeier, F. (2019). Reorganization of spatial configurations in visual working memory. *Memory & Cognition*, *47*, 1469–1480. https://doi.org/10.3758/s13421-019-00944-2

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. II. *Psychologische Forschung*, *4*, 301–350. https://doi.org/10.1007/BF00410640

Wickelmaier, F. (2020). *eba: Elimination-by-Aspects Models* (1.10-0) [Computer software]. https://CRAN.R-project.org/package=eba

Wickelmaier, F., & Schmid, C. (2004). A Matlab function to estimate choice model parameters from paired-comparison data. *Behavior Research Methods, Instruments, & Computers*, *36*(1), 29–40. https://doi.org/10.3758/BF03195547

Wood, J. N. (2011). A core knowledge architecture of visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 357–381. https://doi.org/10.1037/a0021935

Yang, J., Klippel, A., & Li, R. (2015). The cognition of change: Scaling deformations in mind and spatial theories. *Cartography and Geographic Information Science*, *42*(3), 224–234. https://doi.org/10.1080/15230406.2014.991425

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, *24*, 295–340. https://doi.org/10.1016/0010-0285(92)90010-Y

Zacks, J. M. (2020). Event perception and memory. *Annual Review of Psychology*, *71*, 165–191. https://doi.org/10.1146/annurev-psych-010419-051101

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21. https://doi.org/10.1037/0033-2909.127.1.3

Zelinsky, G. J., & Neider, M. B. (2008). An eye movement analysis of multiple object tracking in a realistic environment. *Visual Cognition*, *16*, 553–566. https://doi.org/10.1080/13506280802000752

Appendix A

Table of Equivalences for Category Names

In order to ease readability and understanding, we changed the names of the motion stories appearing in the original paper of Purcalla Arrufi & Kirsch (2018), which we deemed opaque, into more graspable ones. Below we give the equivalence between original names and names used in this manuscript.

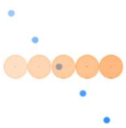
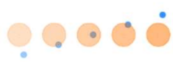
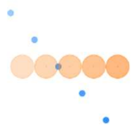
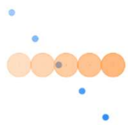
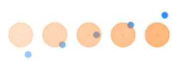
	Story Name	
Motion Categorization	Present Manuscript	Purcalla Arrufi & Kirsch (2018)
Motion-RCC	mr-no	S_{11}
	mr-po	S_{13}
	mr-co	S_{15}
Motion-$OPRA_1$	mo-bp	S_{C21}
	mo-bn	S_{C2-1}
	mo-on	S_{C11}
	mo-op	S_{C1-1}

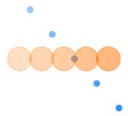
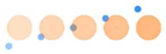
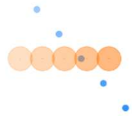
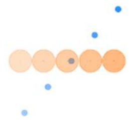
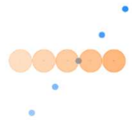
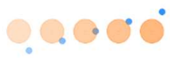
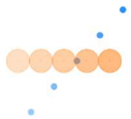
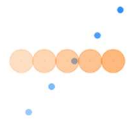
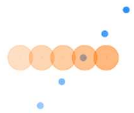
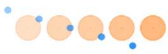
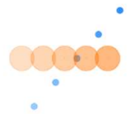
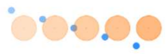
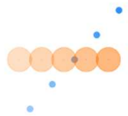
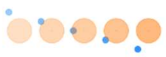
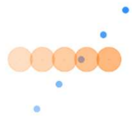
Appendix B

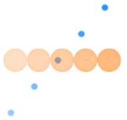
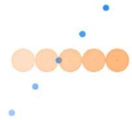
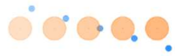
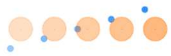
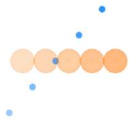
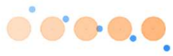
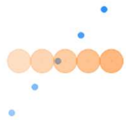
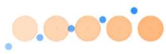
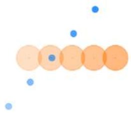
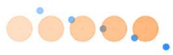
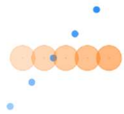
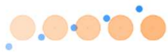
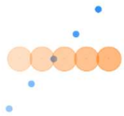
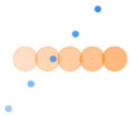
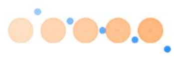
Kinematic Data of the Generated Stimuli

Here we show the complete set of stimuli that we use in our experiments. Each picture depicts five traces of the motion scene, the more faded the discs, the earlier in time the trace is. In the pictures, the bigger entity, o, moves at angle 0°, i.e., towards the positive x-axis – The smaller entity is b. In the experiment, however, we randomly rotated these stimuli for each participant.

The caption at every picture shows the following: the Motion-RCC and Motion-OPRA$_1$ categories (e.g., mr-co mo-bp); and the featural variables d_{min}, α_v, *sgn*(α_{DxDv}). In the featural variables, we can observe how the values of stimulus E are modified symmetrically in the transformed stimuli. For instance, in group Nr. 001, we apply a modification of 1.25 to the reference value d_{min}(E) = 2.75, and we obtain d_{min}(A) = 2.75 – 1.25 = 1.50, d_{min}(C) = 2.75 + 1.25 = 4.00; we apply a modification of 32.08 (rounded value) to α_v(E) = 16.04°, and we obtain α_v(A) = 16.04 + 32.08 ≃ 48.13° (discrepancy due to rounding), α_v(A) = 16.04 – 32.08 = -16.04°

Group Nr.	Reference stimulus	Transformed stimuli			
	E	A RccOpra	B Rcc¬Opra	C ¬RccOpra	D ¬Rcc¬Opra
001	 mr-co mo-bp $d_{\min} = 2.75$ $\alpha_v = 16.04°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-co mo-bp $d_{\min} = 1.50$ $\alpha_v = 48.13°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-co mo-op $d_{\min} = 1.50$ $\alpha_v = -16.04°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-po mo-bp $d_{\min} = 4.00$ $\alpha_v = 48.13°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-po mo-op $d_{\min} = 4.00$ $\alpha_v = -16.04°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$
002	 mr-po mo-bp $d_{\min} = 3.50$ $\alpha_v = 16.04°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-po mo-bp $d_{\min} = 4.50$ $\alpha_v = 48.13°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-po mo-op $d_{\min} = 4.50$ $\alpha_v = -16.04°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-co mo-bp $d_{\min} = 2.50$ $\alpha_v = 48.13°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-co mo-op $d_{\min} = 2.50$ $\alpha_v = -16.04°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$

008	 mr-po mo-bn $d_{\min} = 4.50$ $\alpha_v = 16.04^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-po mo-bn $d_{\min} = 3.50$ $\alpha_v = 48.13^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-po mo-on $d_{\min} = 3.50$ $\alpha_v = -16.04^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-no mo-bn $d_{\min} = 5.50$ $\alpha_v = 48.13^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-no mo-on $d_{\min} = 5.50$ $\alpha_v = -16.04^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$
009	 mr-co mo-op $d_{\min} = 2.75$ $\alpha_v = -16.04^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-co mo-op $d_{\min} = 1.50$ $\alpha_v = -48.13^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-co mo-bp $d_{\min} = 1.50$ $\alpha_v = 16.04^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-po mo-op $d_{\min} = 4.00$ $\alpha_v = -48.13^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-po mo-bp $d_{\min} = 4.00$ $\alpha_v = 16.04^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$
010	 mr-po mo-op $d_{\min} = 3.50$ $\alpha_v = -16.04^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-po mo-op $d_{\min} = 4.50$ $\alpha_v = -48.13^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-po mo-bp $d_{\min} = 4.50$ $\alpha_v = 16.04^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-co mo-op $d_{\min} = 2.50$ $\alpha_v = -48.13^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-co mo-bp $d_{\min} = 2.50$ $\alpha_v = 16.04^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$
011	 mr-no mo-op $d_{\min} = 5.25$ $\alpha_v = -16.04^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-no mo-op $d_{\min} = 6.50$ $\alpha_v = -48.13^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-no mo-bp $d_{\min} = 6.50$ $\alpha_v = 16.04^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-po mo-op $d_{\min} = 4.00$ $\alpha_v = -48.13^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-po mo-bp $d_{\min} = 4.00$ $\alpha_v = 16.04^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$
012	 mr-po mo-op $d_{\min} = 4.50$ $\alpha_v = -16.04^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-po mo-op $d_{\min} = 3.50$ $\alpha_v = -48.13^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-po mo-bp $d_{\min} = 3.50$ $\alpha_v = 16.04^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-no mo-op $d_{\min} = 5.50$ $\alpha_v = -48.13^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$	 mr-no mo-bp $d_{\min} = 5.50$ $\alpha_v = 16.04^\circ$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = +1$

013	 mr-co mo-on $d_{\mathrm{min}} = 2.75$ $\alpha_v = -16.04°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-co mo-on $d_{\mathrm{min}} = 1.50$ $\alpha_v = -48.13°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-co mo-bn $d_{\mathrm{min}} = 1.50$ $\alpha_v = 16.04°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-po mo-on $d_{\mathrm{min}} = 4.00$ $\alpha_v = -48.13°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-po mo-bn $d_{\mathrm{min}} = 4.00$ $\alpha_v = 16.04°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$
014	 mr-po mo-on $d_{\mathrm{min}} = 3.50$ $\alpha_v = -16.04°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-po mo-on $d_{\mathrm{min}} = 4.50$ $\alpha_v = -48.13°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-po mo-bn $d_{\mathrm{min}} = 4.50$ $\alpha_v = 16.04°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-co mo-on $d_{\mathrm{min}} = 2.50$ $\alpha_v = -48.13°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-co mo-bn $d_{\mathrm{min}} = 2.50$ $\alpha_v = 16.04°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$
015	 mr-no mo-on $d_{\mathrm{min}} = 5.25$ $\alpha_v = -16.04°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-no mo-on $d_{\mathrm{min}} = 6.50$ $\alpha_v = -48.13°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-no mo-bn $d_{\mathrm{min}} = 6.50$ $\alpha_v = 16.04°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-po mo-on $d_{\mathrm{min}} = 4.00$ $\alpha_v = -48.13°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-po mo-bn $d_{\mathrm{min}} = 4.00$ $\alpha_v = 16.04°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$
016	 mr-po mo-on $d_{\mathrm{min}} = 4.50$ $\alpha_v = -16.04°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-po mo-on $d_{\mathrm{min}} = 3.50$ $\alpha_v = -48.13°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-po mo-bn $d_{\mathrm{min}} = 3.50$ $\alpha_v = 16.04°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-no mo-on $d_{\mathrm{min}} = 5.50$ $\alpha_v = -48.13°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$	 mr-no mo-bn $d_{\mathrm{min}} = 5.50$ $\alpha_v = 16.04°$ $\mathrm{sgn}(\alpha_{\Delta\vec{x}\Delta\vec{v}}) = -1$