Available online at www.sciencedirect.com
ScienceDirect
Cognition 108 (2008) 533–542
COGNITION
www.elsevier.com/locate/COGNIT

Brief article

When facts go down the rabbit hole: Contrasting features and objecthood as indexes to memory

Merrit A. Hoover *, Daniel C. Richardson

Department of Psychology, University of California, 273 Social Sciences 2, Santa Cruz, CA 95064, USA

Received 14 April 2007; revised 22 February 2008; accepted 29 February 2008

Abstract

People will often look to empty, uninformative locations in the world when trying to recall spoken information. This spatial indexing behaviour occurs when the information had previously been associated with those locations. It remains unclear, however, whether this behaviour is an example of a simple association across perceptual and cognitive systems, or whether location information plays a role in memory retrieval. In the current study, we investigate whether higher-level cognitive processes, such as object-based attention, are involved in spatial indexing. Participants saw creatures burrowing around the screen, appearing from underground to tell them facts. They saw the same creature in two locations, or two identical creatures in two locations, depending on spatiotemporal cues conveyed by a burrowing animation. While answering questions, we found that participants relied on these spatiotemporal cues, fixating the previous locations of the creature associated with the relevant fact, rather than the location of an identical creature. We interpret these findings in terms of an object-based attentional mechanism that is common to semantic memory and scene perception, and allows 'external memory' to be exploited in a dynamic environment.

Keywords: Eye-tracking; Memory; Spatial indexing; External memory; Objecthood

* Corresponding author. Tel.: +1 408 623 4048.
E-mail addresses: mhoover@ucsc.edu (M.A. Hoover), dcr@eyethink.org (D.C. Richardson).

0010-0277/$ - see front matter
doi:10.1016/j.cognition.2008.02.011

1. Introduction

"*It's a poor sort of memory that only works backwards,*" the White Queen remarked – Through the Looking Glass, Carroll (1871).

Alice is surprised that the White Queen claims to remember things from both the past and the future. When Alice replies that she can only remember things that have already happened, the Queen scoffs at this limitation. While we sympathize with Alice, in this paper, we argue that in virtue of the relationship between memory and location, there is a sense in which the Queen is right. We explore whether people associate semantic information with locations based on visual features or objecthood, and argue that the function of this spatial association is a case of memory 'working forwards' as well as backwards.

Spatial information intrudes upon cognitive processes. Simple judgments can be tripped up by conflicts between stimulus and response location (Fitts & Seeger, 1953; Simon, 1969). Encoding of location is thought to occur regardless of the task (Andrade & Meudell, 1993; Chun & Jiang, 1998), and recalling any information about a stimulus recalls information about its location (Brouwer & Van der Heijden, 1997). When a sentence is heard about objects that had previously been present on-screen, people make saccades to the prior location of the relevant object, despite the fact that the objects are no longer present (Altmann, 2004). When objects that are no longer present are described in a narrative, people make eye-movements to the location of the described object (Spivey & Geng, 2001).

Why is location intertwined with these cognitive tasks? Is this evidence of cross-talk between cognitive and perceptual systems, or might location information actually be used for cognitive processing? In this paper, we investigate an interaction between semantic memory, location information and object-based attention. We interpret our results as evidence of spatial information serving a functional role in memory retrieval.

1.1. Spatial indexing

In a semantic memory task, Richardson and Spivey (2000) found that participants were encoding spatial information. Participants heard spoken facts while looking at objects in a two by two grid. These objects were videos of talking heads, in one experiment, and spinning crosses in another. Participants were later asked a question about one of the facts. While answering, they made saccades to the now-empty locations. Even though it was irrelevant to the task, location information was encoded, linked to non-visual information, and activated at retrieval. This 'dynamic spatial indexing' behaviour was found robustly, with different types of visual objects, when grid locations moved between presentation and test, and even when the participants were only six months old (Richardson & Kirkham, 2004).

In a separate paradigm, Altmann and Kamide (2004) found that object locations are tracked and refixated even when the movements are not seen but described linguistically. Participants saw, for example, a glass of wine on the floor and heard, 'the woman will pick up the glass and move it to the table. Then she will pour wine into the glass'. During that second sentence, participants fixated the (empty) location on the table to where the (imaginary) glass had been moved (see also Altmann & Kamide, 2007).

It seems that cognitive processes can bind factual, semantic information to locations in the external world. How are these locations represented? One possibility is that facts are associated with the visual features of objects that co-occur with semantic information. During recall, this visual information is also activated, and a saccade launched to screen locations that look similar (even if those features have moved to a different location). A second possibility is that semantic information is associated with a particular object in the world, which is defined not just by its appearance, but also by its spatiotemporal properties. Such objects would be tracked if they moved around the screen, and re-accessed during recall. In the current experiment, we tease apart these two possibilities, which were conflated in previous work (Richardson & Kirkham, 2004; Richardson & Spivey, 2000). We contrast visual features with object-based encoding in semantic memory.

1.2. Objects and features

Treisman and colleagues (Kahneman, Treisman, & Gibbs, 1992) suggested that object-based information is encoded episodically, and that these episodic representations are updated with each additional view of the object. "Object files" are spatiotemporally contiguous, and are used to facilitate object recognition, among other things. Hommel (2004) extends the application of object files to include "event files" in which object-based mechanisms provide a temporally constraining link between perceptual events, context, and actions performed in that context. These ideas lay the groundwork for examining the role of perceptual encoding in memory tasks.

How are object files linked to objects in the world? Visual features undergo radical changes as a result of lighting shifts, changes in perspective and occlusion by other objects. Scholl (2001) argues that while they play a role in differentiating objects, it is not the visual features per se that are important in defining objecthood. He points to object tracking research which suggests that objects maintain their "objectness" mostly through causal behaviours and spatiotemporal continuity (Pylyshyn & Storm, 1988). In both cases, an object is the same object if it follows realistic behaviour patterns. In order to be perceived as the same object, it must maintain its expected spatiotemporal continuity, behaving appropriately with respect to real world constraints. Does this object-based attentional mechanism play a role in the spatial indexing of semantic memories?

1.3. To what do pointers point?

In our study, spatiotemporal constraints apply to a series of burrowing creatures. Rather than talking heads or spinning crosses (Richardson & Kirkham, 2004; Richardson & Spivey, 2000), participants looked at cartoon animals while they heard pieces of factual information. For example, a white rabbit emerged from a mound of earth in one location while the participants heard a fact about Cleopatra. Later, an identical rabbit emerged from a second mound of earth but no fact was heard. In one condition, an animation of underground burrowing connected these two mounds; in another, the burrowing began from a different, off-screen location. The same object features are seen for the same amount of time across conditions. Spatiotemporal constraints alone imply that in the first case, the same white rabbit is seen in two locations, but in the second case, two visually identical but numerically distinct rabbits are seen. Our central question was: where will participants look when they are asked a question about Cleopatra? If participants ignore the visually identical but numerically distinct rabbit, it will demonstrate that non-visual information is attached to spatiotemporally constrained objects in the world, not simply associated with co-occurring appearances and locations.

2. Method

2.1. Participants

Fourty-two UC Santa Cruz undergraduates participated in partial fulfillment of course requirements. All had normal or corrected-to-normal vision. We could not track 12 participants due to equipment vagaries or certain types of vision correction.

2.2. Apparatus

Participants sat in a cubicle in a reclining chair, looking up at an arm-mounted 19" LCD 60 cm away with a *Bobax3000* remote eye tracker mounted at the base. They could hear stimuli through a headset, and signaled responses with a two-button mouse. An iMac calculated gaze position approximately 30 times a second, presented stimuli and recorded data. Another computer saved an audio-video record of what was seen, heard and said during the experiment, superimposed with gaze position.

2.3. Stimuli

Participants watched a series of animations on a 500×500 pixel brown background subtending approximately 20° visual angle. They saw mounds of earth moving across the screen, as if a creature were burrowing just beneath the surface. Periodically cartoon animals emerged. Participants saw 32 creatures of varying species and colouring, each subtending approximately 3° visual angle. Auditory stimuli were taken from Richardson and Kirkham (2004). The statements either concerned factual information or fictional characters. For each a question was constructed. In half the correct answer was "no" and half "yes".

2.4. Procedure

Fig. 1 shows a schematic of the 16 trials. Each began with burrowing from off-screen which proceeded in a horizontal or vertical direction (2500 ms) ending at

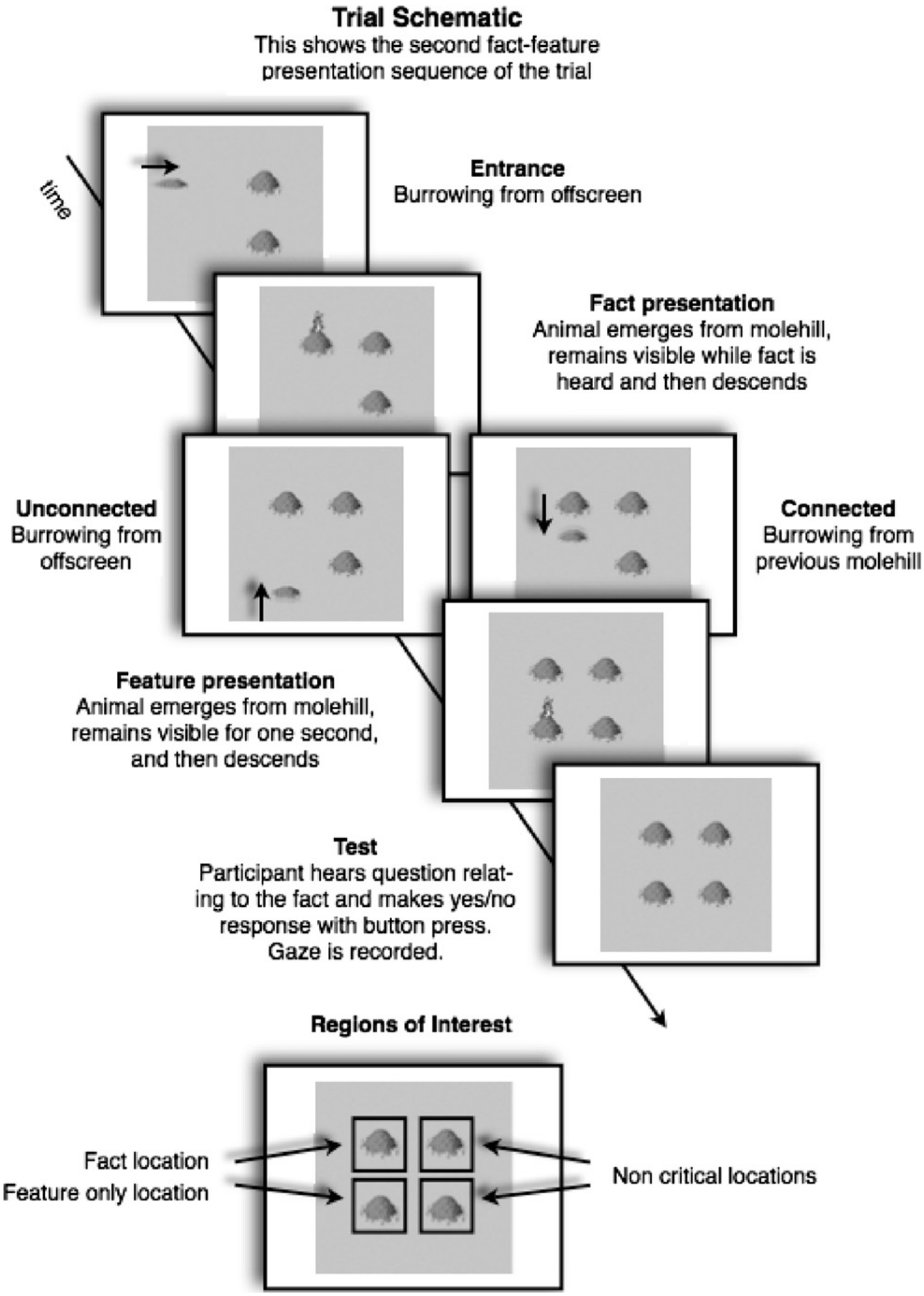

Fig. 1. Trial design and regions of interest.

one of four possible locations in a 2 × 2 grid. An animal then rose up from the hill of earth (500 ms). While the animal was visible, participants heard one of the spoken facts, and then it descended underground. This location was termed the fact and feature location. In the connected condition, participants then saw burrowing (2500 ms) from the last location to a new location. In the unconnected condition, the burrowing ended at the same location but began off-screen. In both conditions, participants then saw a visually identical animal rise from this second, 'feature only' location. In the connected condition, the burrowing implied that this was the same individual; in the unconnected condition it was an identical but distinct animal. In the second appearance, the animal remained on-screen for 1 s in silence, and then descended again, leaving a molehill behind.

The trial continued with a repetition of these steps, so that participants had heard two facts and had seen animals in each of the four locations. Participants then heard a question pertaining to one of the two facts and responded with a button-press. The fact participants were questioned about was termed the 'critical fact', and was associated with two 'critical locations'. The other two locations, associated with the unqueried fact, were termed the 'non-critical' locations. The trial ended with all animals rising from their molehills and scampering off-screen, providing validation of the spatiotemporal cues. There were between two and four animals, depending on how many burrows were connected. The experiment fully randomized and counterbalanced burrowing directions, locations, whether or not the mounds were connected, which question was asked and whether the answer was yes or no. Gaze was recorded from the end of the test question until the participant answered. Regions of interest were defined around the four mounds of earth present on screen during the test question (see Fig. 1).

3. Results

Participants used spatiotemporal continuity as a cue to attach spoken semantic information to locations in the world. Fig. 2 illustrates the timecourse of this behaviour. In the connected condition, where a creature appeared in one location, spoke a fact, then burrowed to another and re-appeared in silence, participants showed an early preference for the feature only locations, but fixated both locations while answering. In the case of the unconnected burrowing, where spatiotemporal constraints implied that two identical but different creatures had appeared, participants focused on the first location of the creature that had been associated with the spoken fact. Throughout the trial, the second location received little more looking than those associated with the non-critical, irrelevant fact.

These conclusions were supported by totaling the looking times in each location during the answer period (Fig. 3) and performing two repeated measures ANOVAs. Firstly, the effect found in previous spatial indexing studies (Richardson & Kirkham, 2004; Richardson & Spivey, 2000) was replicated. Participants looked at screen locations relating to a spoken fact while recalling it. A 2 × 2 repeated measures ANOVA with fact (critical vs non-critical) and burrowing condition (connected and non-con-

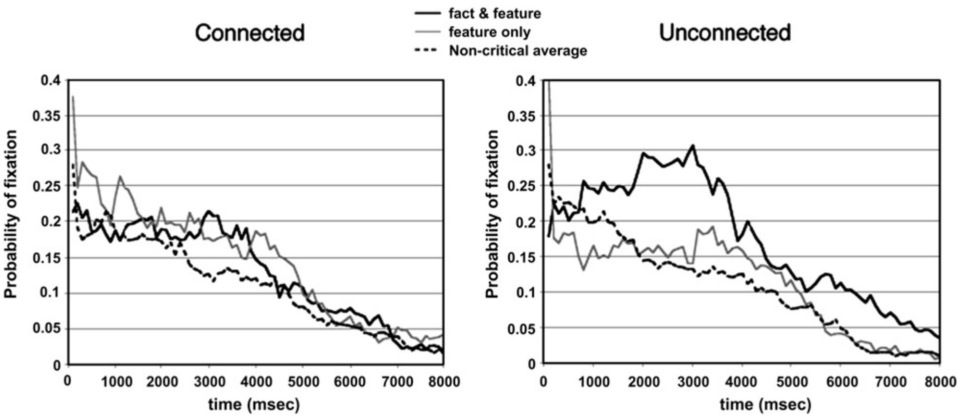

Fig. 2. Fixation probabilities over time for regions of interest in connected and unconnected conditions.

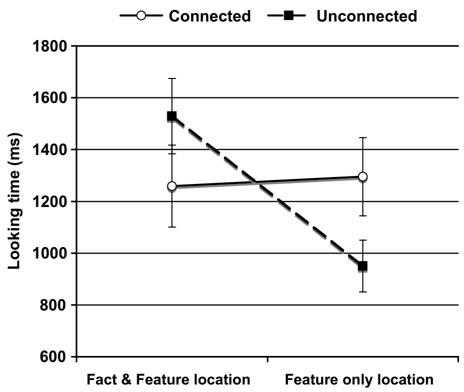

Fig. 3. Mean looking times during answer period. Grey bar represents mean looking times to non-critical locations. Error bars represent standard error.

nected) showed a main effect of fact ($F(1,29) = 18.8$, $p < .001$, ${\eta_p}^2 = 0.38$). Participants looked longer at each of the two locations associated with the critical fact to which the question referred ($M = 1258$ ms), rather than the non-critical locations ($M = 935$ ms). There was no main effect or interaction with the burrowing condition (all Fs < 1).

Having replicated the spatial indexing effect, a second analysis examined how these looks to the two locations associated with the critical fact were affected by the burrowing animations (Fig. 3). A 2 × 2 repeated measures ANOVA with location (fact + features vs features only) and burrowing condition (connected and non-connected) found a significant interaction ($F(1,29) = 4.80$, $p < .05$, ${\eta_p}^2 = 0.14$). There was a main effect of location ($F(1,29) = 6.40$, $p < .05$, ${\eta_p}^2 = 0.18$), but none of the burrowing condition ($F < 1$).

We ran four correlational analyses, examining whether looking times to the two critical locations were related to accuracy in answering the factual questions in the two experimental conditions. All r^2 were less than .01, suggesting that looking at particular location did not predict performance.

4. Discussion

When asked to recall semantic information, our participants attempted to fixate the same object they had seen before, not another object that looked the same. When visual features and spatiotemporal constraints suggested that a single animal had appeared in two of the four locations on screen, participants fixated both while answering. When the spatiotemporal constraints suggested that two identical but different animals had appeared, participants ignored the location of the second appearance. Therefore, in an auditory memory task, spatiotemporal cues are used to narrow down the possible locations that could be associated with the relevant information. In our unconnected condition, these cues override associations with visual features, as there is no increase in looks to the location of a visually identical rabbit. These results show that people use spatiotemporal constraints to track unique objects as they move around the world, and refixate their previous locations when attempting to retrieve non-visual information.

We argue from these results that spatial indexing (Richardson & Kirkham, 2004; Richardson & Spivey, 2000) appears to be object based. But why would an object-based mechanism of visual attention be activated in a semantic memory task? Object tracking is explicitly irrelevant to our task. Indeed looking at certain locations rather than others does not seem to confer a performance advantage. One interpretation of this puzzling behaviour on the part of our participants draws on the notion of 'external memory'. We do not attempt to store an accurate or exhaustive memory of the world, rather we keep in mind only information that is related to the task at hand (Ballard, Hayhoe, Pook, & Rao, 1997). When more information is needed, we use the world as an 'external memory' store (O'Regan, 1992), looking up information in the world rather than looking it up in our memory (Clark & Chalmers, 1998).

In our paradigm, participants heard auditory facts while looking at a burrowing animal on a computer screen. The animal is not explicitly related to the fact, nor is the location on the computer screen. So in what sense can these facts be 'looked up' in the external visual world? There is overwhelming evidence from studies of contextual memory that visual information can provide rich associative cues for semantic information (e.g., Godden & Baddeley, 1975). The participants in our study are not 'looking up' auditory facts in the external world – they are looking up bundles of associated cues that will help them recall the relevant information. Our experiment shows that these bundles are tracked according to their spatiotemporal constraints. This object-based attentional mechanism allows spatial indexing to be useful in a dynamic environment.

Hollingworth (2005) argues that a mechanism of object-based attention stands between the transient perception of a scene and a long-term memory. On the basis of change blindness results, in which large alterations to the external world can go completely unnoticed (Hayhoe, Bensinger, & Ballard, 1998; Rensink, O'Regan, & Clark, 1997; Simons & Levin, 1997), Hollingworth argues that this object-based mechanism consolidates information from visual short-term memory (VSTM) to visual long-term memory (VLTM). What Hollingworth (2005) argues for scene perception, we argue for semantic memory: a mechanism of object-based attention stands between information in the external world and information in memory.

5. Conclusion

It might be the case that we use the world as its own best representation (Brooks, 1991), but every system of information storage needs a system of information retrieval. We claim that when auditory semantic information is encoded, associated visual objects are indexed, tracked using spatiotemporal continuity, and refixated as necessary. This object-based attentional mechanism allows external memory to function in a world where objects move and visual features shift and change (Kahneman et al., 1992). In this sense, the White Queen is right: memory also 'works forward' by indexing sources of contextual information, objects, to be tracked for potential use. By this mechanism, information can reliably be stored and retrieved from external memory.

References

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. *Cognition, 93*, 79–87.

Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57*, 502–518.

Andrade, J., & Meudell, P. (1993). Short report: Is spatial information encoded automatically in memory? *The Quarterly Journal of Experimental Psychology, 46A*, 365–375.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences, 20*, 723–767.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence, 47*, 139–159.

Brouwer, R. T. F., & Van der Heijden, A. H. C. (1997). Identity and position: Dependence originates from independence. *Acta Psychologica, 95*, 215–237.

Carroll, L. (1871). *Through the looking glass, and what Alice found there*. United Kingdom: Macmillan.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*, 28–71.

Clark, A., & Chalmers, D. J. (1998). The extended mind. *Analysis, 58*, 7–19.

Fitts, P. M., & Seeger, C. M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology, 46*, 199–210.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology, 66*, 325–331.

Hayhoe, M. M., Bensinger, D. G., & Ballard, D. H. (1998). Task constraints in visual working memory. *Vision Research, 38*, 125–137.

Hollingworth, A. (2005). The relationship between online visual representation of a scene and long-term scene memory. *Journal of Experimental Psychology: Learning, Memory & Cognition, 31*, 396–411.

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Science, 8*, 494–500.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*, 175–219.

O'Regan, J. K. (1992). Solving the 'real' mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46*, 461–488.

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

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*, 368–373.

Richardson, D. C., & Kirkham, N. Z. (2004). Multimodal events and moving locations: Eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology: General, 133*, 46–62.

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition, 76*, 269–295.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition, 80*, 1–46.

Simon, H. (1969). Reactions towards the source of the stimulation. *Journal of Experimental Psychology*, 81.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1*, 261–267.

Spivey, M. J., & Geng, J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research, 65*, 235–241.