Available online at www.sciencedirect.com
SCIENCE DIRECT®
Robotics and Autonomous Systems 51 (2005) 217–228
Robotics and Autonomous Systems
www.elsevier.com/locate/robot

Erratum

Representation and purposeful autonomous agents☆

Sharon Wood*

Centre for Research in Cognitive Science, Department of Informatics, School of Science and Technology, University of Sussex, Falmer, BN1 9QH, UK

Received 27 July 2004; accepted 27 July 2004
Available online 31 March 2005

Abstract

Although many researchers feel that an autonomous system, capable of behaving appropriately in an uncertain environment, must have an internal representation (world model) of entities, events and situations it perceives in the world, research into active vision, inattentional amnesia has implications for our views on the content of represented knowledge and raises issues concerning coupling knowledge held in the longer term with dynamically perceived sense data. This includes implications for the type of formalisms we employ and for ontology. Importantly, in the case of the latter, evidence for the 'micro-structure' of natural vision indicates that ontological description should perhaps be (task-related) feature-oriented, rather than object-oriented. These issues are discussed in the context of existing work in developing autonomous agents for a simulated driving world. The view is presented that the reliability of represented knowledge guides information seeking and perhaps explains why some things get ignored.

Keywords: Situational modelling; Selective attention; Intent recognition; Active vision; Purposeful autonomous agent

1. Introduction

Traditionally, knowledge representation has been viewed as a prerequisite to informed action. Representations have often been assumed to comprise complete descriptions of the problems solver's environment. Correspondingly, the means by which such representations are constructed and obtain their content are assumed to be comprehensive. For example, early approaches to vision proposed the construction of complete, viewer independent, scene descriptions (cf. [19]).

The development of autonomous agents has changed that view to one which varies from the extreme of denying the existence of representation underlying activity [5,6] to one in which 'representations' of a kind are dynamically generated on a just-in-time (JIT) basis to support interaction with an environment as needed [2,3]. In this view, the purpose of vision is to actively seek out information pertinent to the agent's current

DOIs of original articles:10.1016/j.robot.2004.07.018, 10.1016/j.robot.2005.02.002.
☆ This paper was originally published in *Robotics and Autonomous Systems* 49 (1–2) (2004) 79–90.
* Tel.: +44 1273 678857; fax: +44 1273 671320.
E-mail address: sharonw@cogs.susx.ac.uk.

0921-8890/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.robot.2005.02.003

task, rather than passively absorb information to form a complete 'picture' of the world to be held in memory and interrogated at will in determining appropriate courses of action.

Various psychological evidence appears to support an 'active vision' view. The phenomenon of *inattentional blindness* [11,18,25] or *inattentional amnesia* [25,45] demonstrates the selective nature of vision. Even though entities are clearly within view, if they are not central to the task in hand, they frequently remain unseen [25]. By visually pursuing the selection of information about those entities central to the current task, other entities in the visual scene are actively ignored, no matter how conspicuous they may seem to the non-task-oriented viewer [34].

Guiding this process is the rapid orientation of the viewer to the nature or 'gist' of the situation in which they find themselves and rapid feature selection for task-relevant entities. Gist and spatial layout can be rapidly extracted from visual scenes [4,13,14,36], encoded and retained [9,22], with retention of object identity, recognition, absolute spatial layout, shape, colour, and relative distance occurring gradually over an interval of between 1 and 4 s.

A related phenomenon of *change* blindness further demonstrates aspects of natural vision which result in failure to notice changes to entities in the visual scene when these take place during a saccadic eye movement [10,24]. It appears that change can only be detected when the changing object is fixated [24,27].

This phenomenon has also been demonstrated in a virtual reality setting during activities in which the changed feature is central to the task in hand [11]. Participants asked to pick up blocks, which might be either pink or blue, and to place these in a particular location according to colour, failed to notice when the selected (virtual) object changed colour between initial selection and final placement. Most often the object was placed in the location appropriate for its colour during initial selection, rather than for the colour to which it had changed. Hayhoe [11] argues that this demonstrates the 'micro-structure' of vision: that fixation of an object is not sufficient for apprehension of *all* the visual information associated with it. It would appear that during initial selection of the object, participants pay attention to colour, whilst during subsequent fixations they appear to be concerned with location in guiding the object to its resting place [3]. Furthermore, there is evidence that in natural vision, performance reflects the apprehension of visual information just prior to its use [3].

As a consequence of evidence of this kind, an active vision approach of this kind proposes a task-related basis for the apprehension of visual information. Several models developed using this approach demonstrate that with sufficient sensory input during performance of a task, recourse to internal representation can be avoided [1,2,5,6]. In this sense, just-in-time representation, unlike comprehensive and systematic approaches to representation, appears to refer to currently available sense data, pre-processed to some extent into primitive features, enabling the rapid visual apprehension of task-relevant information.

As a consequence of acquiring task-relevant information, the information itself (such as the colour of a block) may be retained, but the visual context for that information not – that is, just-in-time representation is transitory and merely sufficient for the selection of sought information. There is no enduring representation of the entire scene. It persists only as long as the scene is viewed and is not open to manipulation or reorganization.

What the non-representationalist active vision view appears to be telling us, therefore, is that parsimony in both acquiring information about the world and retaining that information is supportive of appropriate interaction.

The active vision view and computational models based on their approach do not appear to account fully for phenomena observed for natural vision, however. In contrast to the evidence for JIT scene representation, Karn et al. [16] describe the ability to search for or reach towards an object no longer visible. This ability is crucial to many perceptual and motor tasks, and they argue points to the representation of multiple mutually supportive frames of reference for object location. The representation of a viewer independent frame of reference for spatial layout, they state, must be built up over time to support planned activity [12].

This view is not incompatible with notions of JIT scene representation, but it does appear to have implications for retaining some aspects of information about the scenes viewed. The implication appears to be that the rapidly acquired gist and spatial layout of a scene (cf. [36]) is used to support subsequent visual interrogation of a situation. Rather than constructing

a description of all that is in a scene, the findings of Karn et al. [16] and Hayhoe et al. [12] imply that information retained on visual layout enables indexing to the scene to further acquire information as needed. It seems likely that indexing to a situation will be both viewer-centred and perspective independent (cf. [12]), but will not describe the visual information available to the viewer, rather it will support the acquisition of further information as and when required.

Rensink [26] proposes a theory of attention that fits in well with this view. He presents an account of the phenomenon that we experience a rich and detailed account of our visual world, based on being able to actively index visually to the world around us, even though, at any given time, we have access to just that limited set of visual data within our focus of attention. The phenomenon arises from the moment by moment construction of JIT scene representations that give us detailed information about our visual world whenever we want it. The feeling of having pre-stored in some way, the detailed information of all we have ever seen, arises from our ability to relocate our focus of attention to any part of the visual scene, as and when we wish, to 're-discover' all that visual detail in its entirety.

Recent years have seen much progress with saliency-based computational models of attention [15,35] but, as their name suggests, these are primarily responsive to 'attention-grabbing' scene features. There has been some success with using scene-based features associated with context (gist) to further focus active vision processes [20,37,38]. However, there do not currently appear to be any computational models approximating an indexical approach to information acquisition and integration. Evidence for the enactive nature of perception [21] would appear to point to linking an indexical approach to the sensorimotor interactions of an agent with its environment, although exactly how may not be clear cut (cf. [28]). Indeed, Wagner et al. [39] found that particular changes in perspective occur consistently in consequence of particular changes in the viewer's relationship to objects (landmarks) in the scene. The potential for discovering perceived regularities in the environment supporting a viewer-independent interpretation of viewer-dependent sense data offers exciting possibilities in explaining how an index to the environment may itself be derived from information in the environment and thus requiring no representational overhead, and may go some way towards explaining the findings for human vision (cf. [11]).

2. The role of representation

In the context of this view, what then are the implications for the view that an autonomous system must have an internal representation of the events and situations it perceives in the world?

In a dynamic environment which changes outside the individual actions of an agent, that agent is incapable of omniscience with regard to the 'state' of the world; an agent requires access to sense data to support interactions with its environment [41]. Appropriate indexing to just-in-time representations would appear to support this need as such; however, many appropriate interactions with, or responses to, the environment require access to information not available *at the time of response* through sense data alone [42]. In particular, some kinds of activity appear to require anticipation on the part of the viewer, seeming to depend upon the ability to model events involving complex interactions between entities and invoking the application of typical scenario-based knowledge [41,42]. This is a different kind of problem to needing to access information momentarily out of view because the information required is not there to be sensed in the world.

It is unclear to what extent actively indexing to just-in-time scene representations can satisfy certain kinds of anticipatory behaviour. Although there is some evidence that very simple anticipation does not require representation per se [30–32], this has only been demonstrated for cases where the visual information required for generating anticipations (through evocation of learned associations) remains available to the agent's senses throughout. Statistically based predictive properties of such models can be exploited in overcoming the difficulties posed by a limited degree of occlusion of various objects by each other (cf. [29]). However, it is unclear whether these techniques could extend to enable an agent to sustain those anticipations during change in its focus of attention, building up expectations based on the totality of its observations. It seems unlikely, therefore, that JIT representation alone, can provide sufficient basis for all purposeful activity.

The view presented here then is that there is a case for both views; the focus of concern then becomes a matter of integrating JIT scene representation with

more enduring representations of the world and, fundamentally, identifying when it is appropriate to prefer one rather than the other for particular aspects of problem solving.

3. Background

The view presented here draws on earlier work [41] which takes the view that dynamically constructed situational models usefully inform the interactions of autonomous agents in rapidly changing multi-agent domains. This view has largely arisen through the development of a purposeful computer-based agent which carries out the actions of a driver in a simulated driving world [41]. The task domain provides a testbed for investigating goal-directed activity, such as following a route to a particular destination, and the integration of this with dynamically generated responses to situational changes brought about by events, such as changing traffic signals and other agents who may be slowing down for a red light.

A domain such as this imposes demands upon the agent to respond to events in a timely matter. This is achieved through mechanisms for anticipating the outcomes to events. Anticipating outcomes requires the agent not only to go beyond immediate sense data in anticipating how events will proceed but also seems to require the ability to modify expectations in the context of knowledge about the domain (and how agents typically behave within it) combined with evidence of the observed behaviour of other agents, in a given context. The resulting situational world model characterizes the dynamically evolving sequence of events which provide the context for an autonomous agent's activity.

3.1. Agent architecture and situational model

The architecture of the AUTODRIVE system [41] incorporates components which enable the system to interact with a simulated rapidly changing environment. The structure of the agent architecture is characterized in Fig. 1. System components (in square boxes) take inputs, process the input information, and produce outputs (solid lines). Data components (cylinders) store information output by other components. These information stores are accessible by other components (broken lines). For instance, the Sense Data Generator component outputs information about the current driving situation. This is stored as Sense Data where it is accessed by the 'Vision' System component.

Each component plays a role in the overall performance of the system. The role of each component is described briefly followed by further discussion of AUTODRIVE's situational modelling capabilities.

The Sense Data Generator is a microworld simulation program for the domain of driving which provides scenario specific data about what a particular driver can see at a given point in time, rather like a snapshot of the world [40]. The Sense Data contains just that information which is viewable by the driver at a single point in time.

The 'Vision' System has access to this viewable Sense Data. It models attention by selectively viewing objects that the Attention Director has directed it to observe and/or which are in some sense 'attention grabbing', such as sudden changes in the information available. It also models the dual nature of the human visual system by apprehending information spatially locating the agent in its surroundings, for example the vehicle's position in relation to the edge of the highway. The observations made are recorded in Sense Data Memory.

Sense Data Memory stores the results of selective viewing. It contains only a subset of the viewable information – that which has been selectively 'observed'. (Dual visual system processes spatially locating the agent provide an additional data stream to Sense Data Memory, however.) The information stored corresponds to the observation of an object or location at the end point of a single interval of the simulation. A single fixation period may span a number of simulation intervals producing a sequence of data for a particular object or location.

A Situational World Model represents the anticipated *outcome* to observed events based on inputs from Sense Data Memory. The relative distance of fixed features and how these change over time reflects the driver's expectations about his own movements. The anticipated locations of moving objects are based on observations of their current behaviour and expectations about how this might change inferred through Intention Recognition.

Intention Recognition attempts to make meaningful interpretations of observed vehicle behaviour within the situational context in which it takes place, in an at-

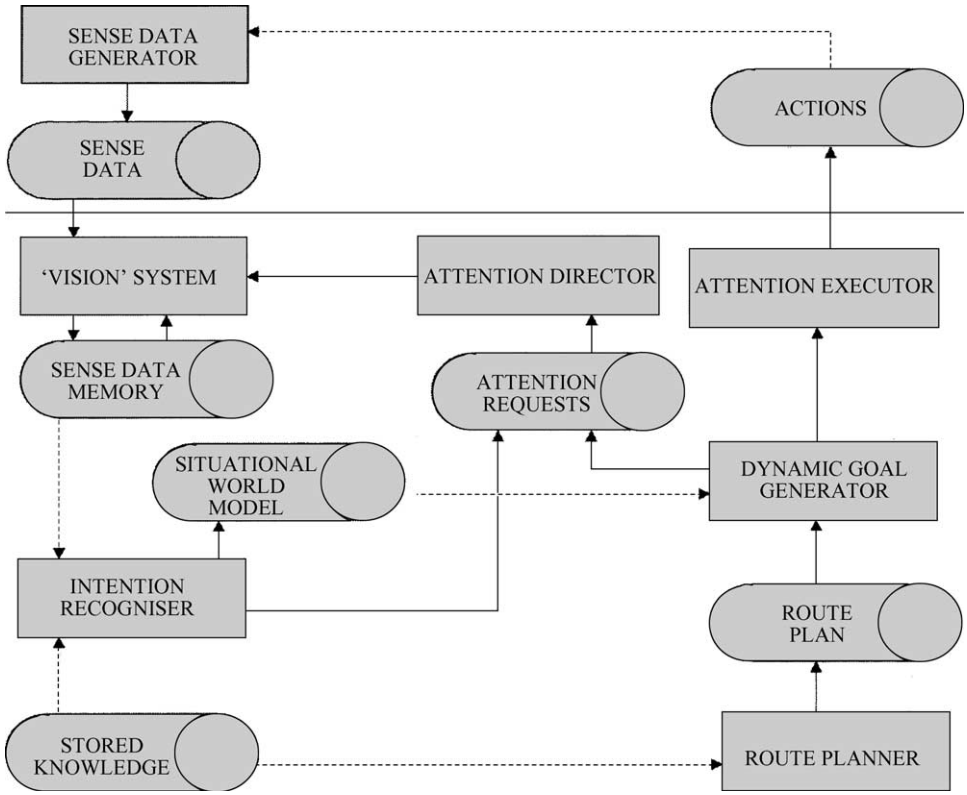

Fig. 1. AUTODRIVE agent architecture.

tempt to hypothesize driver intentions. These hypotheses can provide insight into the future trajectories and velocities of other vehicles in the scene and, in some ways more importantly, identify when *changes* in speed and direction are likely to occur.

Stored Knowledge contains long term information which is not continually updated by new observations. This includes the driver's map-like knowledge about the domain such as the major routes connecting towns and the road networks within towns. On the basis of this the driver is able to identify the route he must follow upon his journey and the turns he must make. (The agent is also able to dynamically follow routes and diversions based on sign information en route.) Stored Knowledge also includes procedurally embedded information about vehicle control, for instance, identifying the speed at which a vehicle may safely take a corner and knowing when to start slowing down. The driver's knowledge about the behaviour of other drivers is similarly founded and provides the basis for identifying constraints on action within specific contexts and hypothesizing intentions.

The Route Planner accesses the route information in Stored Knowledge in formulating a high level Route Plan for reaching the driver's destination. The Dynamic Goal Generator specifies the driver's immediate aims (dynamically generated goals) for realizing the main high level steps of the plan and so ultimately reaching this destination. These immediate aims reflect the time course of constraints on action identified through the Situational World Model. Information to support the generation of appropriate responses is either readily available from the Situational World Model or must be obtained through further observation initiated by specific Attention Requests to the Attention Director.

Requests are sent to the Attention Director in a task-directed manner to obtain information about the world for the purposes of planning and intent recognition. For example, when planning to make a turn, information about oncoming traffic may be sought to supplement

that held in the situational model. The Attention Director prioritises requests in directing the 'Vision' System; only a limited number of observations can be made in the time available for viewing the scene so the Attention Director ensures information required by the Dynamic Goal Generator and Intent Recogniser is weighed up against other important scene features which need attention. Viewing road junctions, for example, may be prioritised when the driver seeks a particular turning, but without sacrificing the need to keep tabs on the movement of other vehicles.

The overall process is one whereby the Situational Model provides interim information whilst the updated results of focused attention are awaited. Similarly, whilst attention is diverted, the situational model maintains the agent's 'awareness' of ongoing events elsewhere in the scene.

The Action Executor translates the immediate aims identified by the Dynamic Goal Generator into brake and accelerator depressions and turns of the steering wheel. These Actions are then relayed to the Sense Data Generator (the world simulator) for their effects to be modelled. The Sense Data generated reflect the driver's altered view of the driving scenario one simulated interval later.

3.2. Situational modelling

The driver may attend to a certain part of the scene over more than one simulation interval or may switch attention to something else. Consequently, a situational world model is constructed incrementally as the agent observes various aspects of its environment. It takes the form of a sequence of snapshots of previously attended objects. Each snapshot provides a collection of viewer-centred descriptions of objects and other agents in the surrounding environment at a single moment in time (rendering continuous processes discrete). The descriptions are based in initial observation of objects and agent behaviour. They are used to predict the consequences of the activities of other agents, and of the viewer's changing perspective on the world. The predictive interval is brief: long enough to inform the viewer's own activities, but necessarily short to reflect the dynamically changing nature of the situation and the extent to which information rapidly becomes out of date. The situational world model enables the viewer to be aware, at any given moment, of its relationship to other agents and objects in the world in which it is situated, even though it may be unable to directly sense those agents and objects at that precise moment in time. Because the situational model is predictive, snapshots of future moments in time enable the viewer to anticipate its future relationships to other agents and objects in the world. A sequence of snapshots therefore describes the way in which the current situation is changing and predicts the outcome to current events.

The situational world model informs the actions of the agent by extending its knowledge of its world into the future. Where the agent's proposed activities conclude some time hence, the world model clearly identifies the constraints on action imposed by future events, enabling the agent to take account of these in realising its goals.

Intent recognition plays a crucial role in informing the accuracy of predicted events. Simple projective anticipation assumes the way the world will change from moment to moment is essentially the same: if another agent is observed travelling at 2 mph over one time interval, simple predictive modelling assumes it will continue to travel at that pace over subsequent time intervals. Intent recognition allows one to insert into that simple calculation the effects of knowledge-based constraints (such as the other agent's inferred goals) that enable prediction of qualitative changes in what happens over one time interval compared to another. For example, in the driving domain, an accelerating vehicle might be expected to cease acceleration when it reaches the speed limit for the highway. Ferguson [8] was able to demonstrate the effect of removing this crucial aspect of anticipation in predicting future events. Using a hybrid, layered agent architecture (consistent with a Brooksian [7] subsumption architecture having no strict hierarchical or prioritised flow of control through layers), it was possible to switch off the intent recognition layer forcing a breakdown in the appropriate timely behaviour of the simulated driving agents.

3.3. A typical scenario

A model of the processes described has been implemented [41,42] and used to simulate various driving scenarios [41] characterizing a range of driving situations. An illustrative example is given below. Each vehicle in the scenario is modelled by a clone of the agent architecture described above. Each vehicle there-

fore receives visual sense data based on its unique position in the scenario corresponding to its own personal perspective. Each vehicle constructs its own personal situational model, attends to whatever is deemed appropriate in achieving its aims, and decides upon its own personal actions. The Sense Data Generator simulates the scenario based on the collective actions of all vehicles at time *t*, generating Sense Data for each cloned agent, viewable by individual drivers at time *t* + 1.

The scenario described here demonstrates the interplay of situational modelling and viewing strategy (as determined by spontaneous and task-directed attentional mechanisms) underlying the ability of a driving agent to modify the behaviour of his vehicle when an obstacle comes into view. The obstacle in this case is a black cat which is crossing the road causing the driver to brake sharply. The ability of the intent recogniser to modify the behaviour of the driver following this car is also demonstrated as he realises that the driver ahead is not behaving as expected. The entire scenario lasts approximately 6 s, reflecting the rapidly changing nature of the situation and interplay of processes for attention and situational modelling. Simulated intervals are set to model 100 ms of real time and the situational model to 8 s.

Time: 0.1–3.6 s. For the purposes of exposition, the scenario involves only two vehicles, a blue car which is positioned 40.0 m along the northern carriageway of a highway and a red car which is 10.0 m behind it at 30.0 m along the same highway. Their initial high level plan step is to traverse the road ahead until the next turning on their route comes into view. Both vehicles begin from a stationary position accelerating towards the speed limit on their way to their destinations in accordance with their immediate aims.

Scene gist (driving domain) is given by the Sense Data Generator which generates task-specific Sense Data only. Spatial layout, situating the vehicle in its physical surroundings, is provided in the Sense Data.

The drivers, as yet, have not begun to construct their situational models of events and so attentional mechanisms are applied on the basis of task only – seeking other vehicles, primarily, and searching the road ahead for obstacles. Spatial layout is apprehended through mimicked parafoveal visual processes. The drivers of the red and blue cars notice each other and proceed to 'fixate' each other over a sequence of 100 ms simulated intervals until sufficient Sense Data Memory has accumulated for them to determine each others' speed and trajectory.

With no evidence to the contrary, each agent's Intent Recognition processes hypothesize a similar intention for each other: to follow the highway whilst accelerating towards the speed limit. Spatial layout indicates the immediate aims of the other driver to achieve this intent within the spatial constraints of the road environment and the context of known obstacles and other vehicles (observed and modelled earlier). As the first objects to be attended, neither of whom at these low speeds poses an obstacle for the other, the drivers situationally model each other pursuing unobstructed acceleration following the known course of the highway.

The driving agents are inhibited from fixating each other again straightaway on the basis that each driver's situational model provides sufficient information about that aspect of the scene for attentional processes to focus upon less central aspects of the scene. In turn, over the space of the next few seconds, the drivers notice a road crossing island, a pair of no-entry signs to a nearby side turning, lane markings and various side turnings as they come into view. The situational model incrementally incorporates each in turn as it is fixated and the relative distance to each object over time is calculated according to the driver's own intended sequence of behaviour.

The overall process is one of an emerging situational model contextualizing new sense data and enabling aspects of what is viewed not only to endure but to contribute to the emerging model of how the situation will change. The model does not describe the scene completely, only attended aspects of it which together constitute an 'awareness' of the situation. So, for example, the relationship over time of the observed vehicle to, say, a specific side-turning would be captured, as would the driver's own relationship to them both.

Owing to the simulated nature of the visual processes, all aspects of the objects viewed are retained in the model (largely positional information, object type (recognition) and identity) rather than selective aspects only (cf. [11]).

Time: 3.6 s. A black cat suddenly appears in the centre of the road presenting an obstacle 55.0 m along the highway. It moves from the centre of the road crossing the drivers' paths to the kerb, a total distance of 6.0 m, at a speed of 3.0 mps. (It is modelled by the Sense Data

Generator and is not a cloned agent.) When deemed viewable by the Sense Data Generator the cat comes into view. The 'cat' appears to the driver of the blue car but is obscured from the red car as it crosses the blue car's path.

Time: 3.8 s. The red car is now 38.11 m along the highway, the blue car, 48.11 m and the black cat 5.4 m from the kerb.

The driver of the blue car is not currently attending to anything in particular owing to inhibitory processes because everything of a task-relevant nature has been observed during the last few seconds. He remains 'aware' of his situational surroundings through his situational model which determines, moment by moment, his immediate aims. The sudden appearance of the attentionally salient black cat elicits a response from the driver becoming the very next thing to be fixated. There is a natural time delay of several milliseconds as the cat is viewed and the consequences of its appearance modelled, and so there is no immediate modification of behaviour at this point. In the meantime, neither driver's expectations about each other have been revised: each expects the other to continue accelerating towards the speed limit and to maintain speed thereafter.

Time: 4.3 s. As the situational model contextualizes this latest observation, anticipating the movement of the black cat, the driver of the blue car ascertains they are on a collision course. He consequently modifies his immediate aims and brakes as hard as possible to avoid collision. (It just so happens the driver of the red car is also braking as he reaches the speed limit and needs to adjust his velocity.) By this time he is less than 5.0 m from the cat with a minimum stopping distance of 5.54 m. There is insufficient space to brake in time to save the cat (in fact, the cat emerged within the minimum braking distance of the blue car making the delay in noticing it irrelevant).

Time: 4.4 s. Attentional inhibition expires and the attention of both drivers switches to each other once again. For the red car, the blue car is a significant obstacle in its pathway; for the blue car, there are no competing vehicles to view so attention naturally falls on the red car as a significant task-relevant entity.

During this time the goals hypothesized for each other are still in effect. The behaviour of the drivers remains unvaried: the blue car is braking hard and the red car continues to drive at the speed limit.

Time: 4.9 s. Until now the driver of the red car has been informed by his situational model-based expectations of the blue car's behaviour as he observes other aspects of the driving situation. Having re-observed the blue car's changing position, he detects a serious discrepancy with his earlier anticipations and so analyses the blue car's behaviour. The overall behaviour of the blue car fails to comply with the immediate aims previously identified and the expectations these gave rise to, and so the red car attempts to identify the other driver's reason(s) for behaving unexpectedly.

The driver of the red car uses his situational model to explore possibilities. This provides immediate information, reflecting his general awareness of the current situation, without the delay of re-observing the surroundings of both agents, various parts of which might now be occluded. He considers possible causes, such as obstacles or vehicles emerging from side turnings, but none are known about. He himself is unaware of the black cat and therefore unable to identify this as the probable cause of the blue car's behaviour. The intention of the driver to make a turn is considered but his behaviour does not appear to comply with executing turns at any of the known side-turnings. In this case, therefore, the driver's behaviour is not consistent with anything that is already known about the prevailing situation.

As the situational model fails to provide evidence of a cause for the blue car's behaviour, so the absence of information initiates visual processes and directs attention to potentially relevant aspects of the scene, for example previously unseen obstacles or side turnings. The situational model guides this process of seeking information in relevant locations in relation to the observed driver. The situational model, therefore, not only identifies what attentional processes should be directed to but also where to seek that information. However, no side-turnings can be found in the vicinity of the blue car's projected position, nor obstacles seen.

Eventually, after failing to identify an alternative cause, the driver of the red car assumes by default the blue car is intending to come to a halt and modifies his expectations accordingly. (The blue car driver's beliefs about the constraints applying to the red car remain unchanged as the red car has given no indication of not conforming to these. However, his reassessment of those constraints in the current context of the situational model include himself as an obstacle to the red car.)

Time: 5.3 s. The driver of the blue car has almost reached the black cat now and continues in his attempt to avoid it by braking as hard as possible; it looks like a lost cause as his minimum stopping distance (2.22 m) exceeds the distance within which he is able to stop (0.35 m). The driver of the red car, aware he is on a collision course with the blue car, is also braking.

Time: 5.5–6.0 s. The driver of the blue car, despite braking as hard as possible, passes the cat. The location of the cat is captured in the driver's situational model; his awareness of this change in constraints on his action, enables him to resume his immediate aim to accelerate towards his destination. (Were he to look in his rear view mirror, perhaps he would see the black cat, having made a last minute dash, narrowly escape his wheels.)

The driver of the red car, however, has no such change in expectations based on his own situational model; he continues to expect the blue car to stop, and therefore continues to brake. He will not alter his expectations until he re-observes the blue car a few seconds later and detects the change in circumstances.

3.4. Limitations

Overall the scenario is interesting in the way we see a breakdown of the intent recognition processes in a situation where the unexpected and unpredictable takes place; such an event enables the demonstration of the interplay between situational modelling and the agent's viewing strategy as an adjunct to spontaneous and otherwise more straightforward task-directed attentional processes. The scenario provides a good demonstration of the value of situational modelling. Other entities encountered in the scenario contribute to the agent's awareness of the situation, and the modelling of these is supportive of the agent's ability to direct attention to that aspect of the scene that most demands it: the blue car. The situational model, in rendering other aspects of the scene predictable, assists the agent in addressing the unpredictable. Furthermore, it is the discrepancy between earlier predicted behaviour and subsequent observations that highlights the need for attention to be focused, perhaps more than otherwise, on understanding the behaviour of the blue car.

Attentional inhibition in the original model of attention [41] is relatively inflexible and further development of the model involved a more dynamic, flexible, probabilistically guided approach [44]. However, the model implemented is capable only of mimicking the attentional processes evident in viewing natural scenes. Although conceptually compatible with JIT models of scene interpretation, AUTODRIVE's model of attention is designed to work with the output of a simulated world. A model of attention showcasing the partnership between JIT scene representation and situational modelling would be more appropriately based on the processing of natural image data streams to demonstrate the role of situational modelling in overcoming sensing limitations and supporting problem solving based on anticipation of events not yet taken place. A model of this kind may provide insight into the successful indexing of attention to relevant parts of the scene in seeking information, of a kind afforded in the scenario above. This is the focus of current work.

3.5. Integrity and completeness of the situational model

As the scenario helps to demonstrate, situational modelling provides an informed context for analyzing current sense data, and a platform for realizing its implications for future scenarios [41–44]. Of course, the advantages gained through situational modelling depend upon the quality of the knowledge held in the model. As a model of current events based in previously sensed data, its validity diminishes over time. Uncertainty in the sensed data and its implications contributes to this effect. Therefore, objects and agents remaining within view must be repeatedly observed if the integrity of the situational model is to be maintained.

Typically the viewer is limited in the rate at which it can apprehend new information. Consequently the model cannot provide a 'complete' description of the agent environment, rather it is selective in the information described. The incompleteness of the model upon which an agent relies in order to interact effectively with its environment is not inherently problematic, provided the model is good enough [33,44]. Indeed, the findings on inattentional amnesia and change blindness clearly support this view. It is the inherent 'goodness' or integrity of the model which can guide us here: rather than seeking 'information about the world' an agent might seek 'information to maintain the integrity of its situational model', as it relies upon this to inform action. Identification of lapses in the integrity of knowl-

edge held in the model can be used to guide sensing priorities, informing focus of attention and selective perception, with the aim of maintaining the quality of knowledge held [44].

4. To represent or not to represent is not the question

Viewing the situational model as not only a means of informing action but also of informing viewing strategy would also appear to offer a basis for understanding perhaps why some things get ignored.

We have already considered the phenomenon of 'change blindness' when a change in a central aspect of a scene remains undetected owing to attentional limitations [24]. What is intriguing is that a similar phenomenon is observed even though the changed feature is central to the agent's activity, as in Mary Hayhoe's [11] experiments. Participants behave in accordance with having apprehended the colour of the objects to be manipulated, so this information would appear to be held in memory; why then failure to observe subsequent changes to this feature?

Hayhoe (*in conversation*) made a further observation that on occasions when participants *did* notice the new colour of the held object, they would frequently conclude they had mistakenly picked up another object to that intended, rather than that the object's colour had changed (even though this possibility had been mentioned in instructions to participants).

Both phenomena described: (i) failure to interrogate the visual scene for object colour following its initial designation, and, (ii) when the object *was* re-interrogated for colour, the assumption that colour change was a result of misperception rather than actual change, could be explained through a particular characteristic of visual information seeking. It would seems that knowledge about our visual world, learned through experience, tells us that certain aspects of a visual scene are more enduring than others. Objects rarely change colour; following initial identification, therefore, there is little reason for checking an object's colour again. Consequently, a change in colour is more attributable to an error in its initial perception or, more likely still in a cluttered scene, the failure to direct an action towards the intended object. Experience is a powerful determinant of visual experience [17,23] and it would seem to be at least possible that experience might also play this role in guiding viewing strategy (cf. [26]). Expectations regarding persistence effects would point to it being safe to ignore some aspects of our world over others, once the crucial information required has been initially apprehended.

This alternative interpretation of the data indicates the possibility that the evidence for 'change blindness' may not always point to a failure to represent sensed data, but rather a failure to question the validity of sensed data when used in later problem-solving and in informing action which then results in ignoring subsequent change through failure to reinterrogate the scene for validation.

The implication of this view is that sensing the world is primarily an information-maintenance activity, rather than an information-discovery activity. Agents sense, but ignore not only what they don't need to know, but what they think they already know, choosing to notice not what needs to be noticed but what is believed to be unknown.

The observation, if correct, that visual behaviour is consistent with experience, might also guide us in determining where and when an autonomous agent would be best advised to situationally model, and when not: one size need not necessarily fit all – model where experience indicates modelling works, but do not model for situations where there is no information advantage in modelling; or, where experience tells us the situation is difficult to anticipate, adopt alternative tactics such as combining minimal prediction with frequent monitoring instead.

5. Summary

Purposeful autonomous agents would seem to require the ability to model the world around them if they are to be able to interact effectively with that world. On the other hand, evidence from human studies suggests that much activity can be supported through JIT scene representation. This poses the new problem of how to integrate in a coherent and useful way, knowledge reflecting situational understanding stored in memory, with transitory information from our visual surroundings. Ontologically, the knowledge representation techniques used should support the natural task-based feature-oriented way in which the

environment is interrogated and reasoned about in problem solving activities.

The problem of identifying where and when it is preferable to construct more enduring representations, and the problem of identifying when to update them, appear to be related. It is suggested that evidence from human studies on the relationship between learned expectations and the likelihood that change will be ignored, could guide further investigation of these difficult but interesting and challenging problems.

References

[1] R. Bacjsy, Active perception vs. passive perception, in: Proceedings of the Workshop on Computer Vision, vol. 5, 1985, pp. 55–59.

[2] D. Ballard, Animate vision: an evolutionary step in computational vision, J. Inst. Electron. Info. Commun. Eng. 74 (1991) 343–348.

[3] D. Ballard, M. Hayhoe, J. Pelz, Memory representations in natural tasks, J. Cogn. Neurosci. 7 (1995) 66–80.

[4] I. Biederman, On the semantics of a glance at a scene, in: M. Kubovy, J.R. Pomerantz (Eds.), Perceptual Organisation, Lawrence Erlbaum Associates, Hillsdale, NJ, 1981.

[5] R. Brooks, A robust layered control system for a mobile robot, IEEE J. Robot. Auto. 2 (1986) 14–22.

[6] R. Brooks, Intelligence without representation, Artif. Intel. 47 (1991) 39–59.

[7] R. Brooks, How to build complete creatures rather than isolated cognitive simulators, in: K. Van Lehn (Ed.), Architectures for Intelligence, Lawrence Erlbaum Associates, Hillsdale, NJ, 1991.

[8] I. Ferguson, TouringMachines: An Architecture for Dynamic, Rational, Mobile Agents, TR Report No. 273, Computer Laboratory, University of Cambridge, Cambridge, 1992.

[9] A. Friedman, Framing pictures: the role of knowledge in automatized encoding and memory for gist, J. Exp. Psychol.: Gen. 108 (1979) 316–355.

[10] J. Grimes, On the failure to detect changes in scenes across saccades, in: K.A. Atkins (Ed.), Perception, Vancouver Studies in Cognitive Science, 5, OUP, Oxford, 1996.

[11] M.M. Hayhoe, What guides attentional selection in natural environments? in: Abstract Proceedings of the Fifth Workshop on Active Vision, University of Sussex, September 22, 2003.

[12] M.M. Hayhoe, A. Shrivastava, R. Mruczek, J.B. Pelz, Visual memory and motor planning in a natural task J. Vision 3 (2003) 49–63.

[13] H. Intraub, Presentation rate and the representation of briefly glimpsed pictures in memory, J. Exp. Psychol.: Human Learning and Memory 6 (1980) 1–12.

[14] H. Intraub, Rapid conceptual identification of sequentially presented pictures, J. Exp. Psychol.: Human Perception and Performance 7 (1981) 604–610.

[15] L. Itti, C. Koch, E. Neibur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Machine Intel. 20 (11) (1998) 1254–1259.

[16] K.S. Karn, P. Moller, M.M. Hayhoe, Reference frames in saccadic targeting, Exp. Brain Res. 115 (1997) 267–282.

[17] R.B. Lotto, D. Purves, The empirical basis of colour perception, Conscious. Cognition 11 (2002) 609–629.

[18] A. Mack, I. Rock, Inattentional Blindness, MIT Press, Cambridge, MA, 1998.

[19] D. Marr, Vision, W.H. Freeman, Oxford, 1982.

[20] A. Oliva, A. Torralba, Modeling the shape of the scene: a holistic representation of the spatial envelope, Int. J. Comp. Vision 42 (3) (2001) 145–175.

[21] J.K. O'Regan, A. Noe, A sensorimotor account of vision and visual consciousness, Behav. Brain Sci. 24 (2001) 939–1031.

[22] K. Pezdek, T. Whetstone, K. Reynolds, N. Askari, T. Dougherty, Memory for real-world scenes – the role of consistency with schema expectation, J. Exp. Psychol.: Learning, Memory and Cognition 15 (1989) 587–595.

[23] D. Purves, R.B. Lotto, S. Nundy, From the cover: why we see what we do? Am. Scientist 90 (2002) 236–243.

[24] R.A. Rensink, Change detection, Ann. Rev. Psychol. 53 (2002) 245–277.

[25] R.A. Rensink, Seeing, sensing and scrutinizing, Vision Res. 40 (2000) 1469–1487.

[26] R.A. Rensink, The dynamic representation of scenes, Visual Cognition 7 (2000) 17–42.

[27] R.A. Rensink, J.K. O'Regan, J.J. Clark, On the failure to detect changes in scenes across brief interruptions, Visual Cognition 7 (2000) 127–145.

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

[29] K. Sage, A.J. Howell, H. Buxton, Developing context sensitive hmm gesture recognition, in: Proceedings of the Fifth International Workshop on Gesture and Sign Language Based Human-Computer Interaction, Genova, Italy, April, 2003, pp. 277–287.

[30] M. Schlesinger, A. Barto, Optimal control methods for simulating the perception of causality in young infants, in: M. Hahn, S.C. Stones (Eds.), Proceedings of the 21st Conference of the Cognitive Science Society, Erlbaum, 1999.

[31] M. Schlesinger, P. Casey, Visual expectation in infants: evaluating the gaze-direction model, in: Boston College, Boston, MA, August 4–5, Proceedings of the Third International Workshop on Epigenetic Robotics, 2003, pp. 115–122.

[32] M. Schlesinger, D. Parisi, The agent-based approach: a new direction for computational models of development, Dev. Rev. 21 (2001) 121–146.

[33] H.A. Simon, The Sciences of the Artificial, MIT Press, Cambridge, MA, 1981.

[34] D.J. Simons, C.F. Chabris, Gorillas in our midst: sustained inattentional blindness for dynamic events, Perception 28 (1999) 1059–1074.

[35] Y. Sun, R. Fisher, Hierarchical selectivity for object-based visual attention, in: Proceedings of the Second Workshop on Biologically Motivated Computer Vision, Tubingen, Germany, November, 2002.

[36] B. Tatler, I. Gilchrist, J. Rusted, The time course of abstract visual representation, Perception 32 (2003) 579–592.
[37] A. Torralba, Modeling global scene factors in attention, J. Opt. Soc. Am. 20 (7) (2003) 1407–1418.
[38] A. Torralba, Contextual priming for object detection, Int. J. Comp. Vision 53 (2003) 169–191.
[39] T. Wagner, U. Visser, O. Herzog, Egocentric qualitative spatial knowledge representation for physical robots, in: Proceedings of the AAAI Spring Symposium on Knowledge Representation and Ontology for Autonomous Systems, March, 2004, pp. 9–16.
[40] S. Wood, Dynamic world simulation for planning with multiple agents, in: Proceedings of the Eighth International Joint Conference on Artificial Intelligence (IJCAI-8), Karlsruhe, W. Germany, August 8–12, 1983, pp. 69–71.
[41] S. Wood, Planning and decision-making in dynamic domains, Ellis Horwood, Chichester, 1993.
[42] S. Wood, When being reactive just won't do, in: Proceedings of the AAAI Spring Symposium on Integrated Planning Applications, Stanford University, Palo Alto, CA, March 27–29, 1995, pp. 102–106.
[43] S. Wood, The role of plan recognition in reducing situational uncertainty and directing attention in planning, in: Proceedings of the IJCAI Workshop on The Next Generation of Plan Recognition Systems: Challenges for and Insight from Related Areas of AI, Montreal, Canada, August 20, 1995, pp. 124–129.
[44] S. Wood, A Probabilistic Network Approach to Prioritising Sensing and Reasoning: a step towards satisficing modeling, in: Proceedings of the AAAI Spring Symposium on Satisficing Models, Stanford University, Palo Alto, CA, March 23–25, 1998, pp. 87–90.
[45] J.M. Woolfe, Inattentional amnesia, in: V. Coltheart (Ed.), Fleeting Memories, MIT Press, Cambridge, MA, 1999.

Sharon Wood received her BAHons. degree in Developmental Psychology with Cognitive Studies, and her DPhil degree in Computer Science and Artificial Intelligence from the University of Sussex (England) in 1979 and 1990, respectively. Between 1983 and 1987 she held various research posts at the University of Bath (England) and University of Sussex. Since then she has been a Lecturer in Computer Science and Artificial Intelligence in the Department of Informatics (formerly School of Cognitive and Computing Sciences) University of Sussex.