Cognitive Systems Research 95 (2026) 101420

Contents lists available at ScienceDirect

Cognitive Systems Research

journal homepage: www.elsevier.com/locate/cogsys

Generating models of attentional cueing and inhibition of return with genetic programming

Laura K. Bartlett [a,*], Noman Javed [a], Dmitry Bennett [a], Peter C.R. Lane [b], Fernand Gobet [a]

[a] The Centre for Philosophy of Natural and Social Science, London School of Economics and Political Science, UK
[b] School of Physics, Engineering and Computer Science, University of Hertfordshire, UK

ARTICLE INFO

Keywords:
Genetic programming
Attention
Cueing
Inhibition of return

ABSTRACT

The cueing task is a robust experimental paradigm for investigating attention. A centrally presented valid cue, correctly indicating the location of an upcoming target stimulus, leads to quicker responses than an invalid cue. A feature of this paradigm is that increasing the delay between a peripheral cue and a target reverses this effect, where responses become slower for a valid cue, a phenomenon termed inhibition of return (IOR). Using GEMS, a system that utilises genetic programming techniques, we generated potential strategies underlying the facilitation and IOR effects in the cueing paradigm. Models were generated for three experiments differing in their experimental designs, all with good fit to behavioural data. Our approach helps address current issues in the field of attention regarding how it is defined and what mechanisms underlie it. Additional benefits and limitations of this method are discussed.

1. Introduction

Attention has been a central concept in the study of human cognition for decades. It is often defined as the selective processing of features of the environment, and inhibition of irrelevant sensory information (Carrasco, 2011). Attention is often considered a necessary mechanism due to limited cognitive and neural resources available for processing the wealth of sensory information coming from the environment. It essentially allows us to ignore features of the environment that are not important, and enhance those features that are task-relevant, novel or salient.

Different kinds of attention have been described, often as dichotomous concepts. For example, attention that requires eye movements is overt, while covert attention refers to shifts in attention without eye movements. Similarly, attention can be described as endogenous, where focus is deliberately shifted, or exogenous, where attention is reflexively captured by something in the environment. Often, these concepts are teased apart with the cueing experiment (Posner, 1980), which involves an initial cue, followed by a target stimulus either in the cued or uncued location (see Fig. 1 for an example trial). This simple and popular paradigm is often used to investigate aspects of selective attention. For example, comparing a central arrow cue, which requires deliberate moving of attention, against a sudden-onset peripheral cue that catches attention, has been suggested to give insight into exogenous and endogenous attention.

While attention is often taken for granted in the literature, a comprehensive understanding of its components and mechanisms remains largely absent and the use of the concept itself has been criticised. For example, Anderson (2011) argued that constructing and testing such dichotomies has substantially slowed progress in understanding attentional effects, and has only worked to show that such dichotomies do not exist. Indeed, an uninformative arrow cue will still affect behaviour (Ristic, Friesen, & Kingstone, 2002), suggesting that both central and peripheral cues can capture attention. There are also issues with definitions – attention can be described in many different ways, often without a clear conceptualisation, and often definitions are informal and verbal. Frequently, attention is used as a placeholder for 'something' rather than a detailed theoretical concept (Anderson, 2011). While some researchers argue that the term has become unproductive and should be abandoned due to the range and diversity of meanings and confusion over what attention is (Anderson, 2021; Hommel et al., 2019), others argue that attention could still be a useful concept (Taylor, 2023). Throughout years of study and experimentation, there are now a number of unwieldy theories that, instead of being disregarded or overhauled, often tack on new features to explain new results. This results in increasingly complex

* Corresponding author.
E-mail addresses: l.bartlett@lse.ac.uk (L.K. Bartlett), n.javed3@lse.ac.uk (N. Javed), d.bennett5@lse.ac.uk (D. Bennett), p.c.lane@herts.ac.uk (P.C.R. Lane), F.Gobet@lse.ac.uk (F. Gobet).

https://doi.org/10.1016/j.cogsys.2025.101420
Received 1 December 2024; Received in revised form 9 September 2025; Accepted 28 October 2025
Available online 25 November 2025

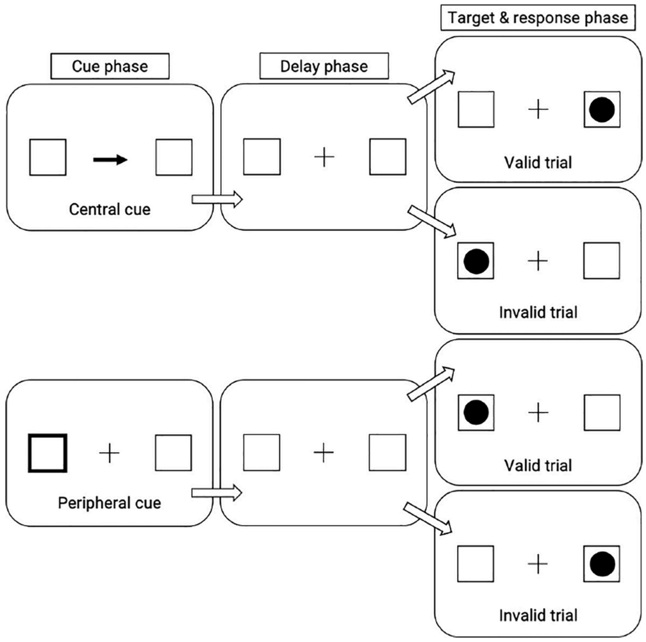

Fig. 1. Example trials of the central cueing (upper) and peripheral cueing (lower) experiments.

theories with potentially ill-defined additional features, and reduced predictive capabilities. In a similar vein, it has been argued that there is a lack of theory-driven experimentation or interpretation of data (Anderson, 2021). In order to constrain definitions and theories of attention, researchers have been demonstrating that effects often attributed to selective attention could instead be accounted for by other factors, such as efficient peripheral vision encoding (Rosenholtz, 2018).

Given these criticisms, researchers must scrutinize definitions of attention more thoroughly. However, use of the term is pervasive, and it is within the context of previous research and theories that we interpret and understand research findings. Researcher bias is a strong factor in not only how results are interpreted, but which experiments are planned and how they are designed. The growing discontent within psychology due to many years of often verbal theories that lack precision, and a lack of coherence across different psychological domains contribute to issues such as the replication crisis (Open Science Collaboration, 2015). As such, new techniques need to be developed.

In the current paper, we show how the Genetically Evolving Models in Science (GEMS) methodology (Bartlett et al., 2023b,c), which uses the artificial intelligence (AI) technique of genetic programming (GP), can construct novel and well specified models of attention that can account for the human data. Based on the models generated, new experiments can be designed to test the models and their predictions.

2. The cueing paradigm

As described above, a particularly influential paradigm in the attention literature is the cueing task (see Fig. 1) – effects are generally robust across experiments, and the paradigm has been consistently used since its popularisation by Posner (1980). The cueing paradigm is ideal for computational modelling, as the effects are well established, consistent and reliable. Further, the simplicity and flexibility of the experiment has resulted in a wealth of data with variations in conditions, timings, and participant pools.

As demonstrated in Fig. 1, the basic task consists of an initial cueing phase, followed by the presentation of a target stimulus, which participants must respond to. For this paradigm, a number of design decisions are made based on what is being investigated and the goals of the experiment (see Chica et al. (2014) for a detailed overview of the task and the variables that can be adjusted). Typically, a central arrow cue that correctly indicates the location of the target (a valid cue) results in quicker and more accurate responses compared to an invalid cue, where the central arrow is not directed towards the target. Details of this facilitation effect have been thoroughly explored: it is seen for unpredictive cues (e.g., for experiments with 2 possible target locations, the cue is correct 50 % of the time), increases with cue predictability (Arjona et al., 2016), typically lasts for a number of seconds (Posner, 1980) and occurs when target onset is approximately 300 milliseconds after cue onset (stimulus onset asynchrony (SOA); Remington and Pierce (1984)).

A number of computational models have been developed to explore attentional processes, both symbolic and connectionist. For example, the symbolic cognitive modelling architecture of ACT-R (Anderson et al., 2004) involves components relating to attention. These components have been extended to account for more attentional processes. For instance, ACT-R has been used to model the attentional network test (ANT) developed by Fan et al. (2002). This task is a variation of the cueing task described in the current paper, consisting of a cueing phase, followed by a test phase that presents the Flanker task (Eriksen & Eriksen, 1974) at either the cued or uncued location. The flanker task has a target (typically an arrow) at the centre of the display, with 2 arrows on either side that either point in the same direction or opposite direction to the central target. This task was designed to test a theory of attention comprising of 3 elements: alerting, orienting and executive control (Posner & Petersen, 1990). Behavioural results of the ANT were well fit by a modified version of ACT-R (Hussain & Wood, 2009; Wang et al., 2004).

The cueing experiment has also been modelled with the connectionist model of MORSEL (multiple object recognition and attention selection; Mozer (1991), Mozer and Sitton (1998)). This model can simulate visual search data, and demonstrate generalisability to the Posner cueing task. It has two components: an object recognition system and an attentional mechanism that modulates levels of activation in the recognition module. More specifically, activation levels are raised for stimuli in attended areas compared to unattended areas, similar to a spotlight conceptualisation of attention (Posner et al., 1980). This spotlight is not fixed though, with excitatory local connections and inhibitory connections for distant units, akin to the zoom lens model of attention (Eriksen & St. James, 1986). For the cueing task, target location was pre-activated by the cue, which reduced the amount of activation needed once the target appeared in the cued location, but caused competition following invalid cues.

These methods of modelling attentional cueing effects are important for furthering our understanding of the mechanisms underlying attention. However, GEMS provides another angle for exploring potential theories. In particular, rather than creating a theory and testing how well it fits experimental data, GEMS can include operators (i.e., basic mechanisms) from multiple different theories and from different psychological domains, and evolve models that best fit the data.

While the cueing paradigm is often used to explore the facilitatory effect of a valid cue, it also serves to explore the inhibitory effect of a valid cue, termed inhibition of return (IOR). This phenomenon occurs following a peripheral cue at a particular location, where responses are slower for target stimuli that are presented at the cued location. This leads to the pattern of results shown in Fig. 3: while at short interstimulus intervals (ISI) a valid cue facilitates responses, when the interval is longer a valid cue increases response times.

IOR is a surprising effect, given the robust findings for reduced errors and reaction times for valid, compared to invalid, cues. In one of the initial demonstrations of the effect, Posner and Cohen (1984) suggested that IOR might occur because of novelty seeking. Inhibition has also been suggested to be functional for search by moving away from areas already inspected, and as such useful for foraging (Klein, 2000).

The mechanisms and circumstances under which IOR can be found have been extensively studied (for a review, see Lupianez et al., 2006). IOR is only present for peripheral cues, and occurs even when participants are aware that the cue is not informative. The timecourse of the effect has also been explored; IOR lasts for over three seconds (Samuel & Kat, 2003), and can occur for ISIs from 250 ms (Klein, 2000; Samuel & Kat, 2003), though more demanding tasks have a later onset IOR (Lupianez et al., 1997). The link between inhibition shown in the cueing paradigm and eye movements was made by Posner et al. (1985). When participants respond by moving their eyes towards the stimulus that they perceived first, these movements were more often made in the direction away from the cued location, suggesting that eye movements were biased by the cue. Behavioural results may also be accounted for by sensory adaptation mechanisms, where sensitivity is reduced following exposure to a stimulus (Hilchey et al., 2014; Lim et al., 2018).

The effect of IOR, despite the interest of psychologists and numerous experiments, is still not consistently understood. Dukewich and Klein (2015) explored how IOR assumptions varied across experts, finding unique points of view regarding the phenomenon and concluding that a coherent theory was lacking. There are some concerns regarding the overuse of the term, and differences in experimental conditions could lead to different underlying mechanisms being employed. This relates to the issues outlined regarding attention research more generally. Computational modelling with comprehensive theories are beneficial for understanding IOR (Satel et al., 2019). We demonstrate the potential of symbolic modelling to address some of the issues raised, with a view to better understanding attention and IOR.

The simplicity and flexibility of the cueing paradigm make it highly informative about cognitive processes. The possibilities that the flexibility of the paradigm allows has, however, resulted in a great number of experiments with slightly different timings and stimulus positions without theoretical justification. As such, it can be difficult to have a clear picture of the mechanisms underlying the paradigm. For example, comparing experiments with different stimulus display durations might give insight into the effect of stimulus duration, but if other factors (e.g., distance from fixation, method of response, etc.) are arbitrarily different, then clear conclusions cannot be made.

2.1. Automating scientific theory development

One viable direction for the field is to harness the ever increasing power of AI. With continuous improvements of both hardware and software, AI has become more integrated into scientific practice (Musslick et al., 2025). It has already been embraced in fields that have precise goals and specific procedures (e.g., functional genomics, King et al., 2004). AI techniques have similarly been gaining interest and use for scientific discovery in psychology. For example, Peterson et al. (2021) used machine learning to develop theories of decision-making. For an overview of computational modelling and AI in psychology, see Bartlett et al. (2023a).

In this paper, we generate a number of new models of attention using GEMS to better understand patterns of cue facilitation and IOR in the peripheral and central cueing paradigms. This technique generates candidate models and evolves them over many generations to find those with the best fit according to a specified fitness function. First generation models are created by randomly sampling operators, and those with the best fit move into the next generation. A selection of these models undergo alterations, namely crossover (swapping sections of models) and mutation (changing an operator of a model for something new). A notable feature of GP is that entire programs are evolved and can vary in size, rather than the fixed length strings characteristic of genetic algorithms.

Reinforcement learning (RL) is another technique that has been employed to sequence actions with the aim of maximising a fitness function or a reward signal. A comparison between the two approaches is therefore appropriate at this stage. Both genetic programming (GP) and reinforcement learning (RL) optimise behaviour by searching for solutions that maximise a performance criterion, but they differ in how they explore and update candidates. As just noted, GP evolves a population of candidate programs using selection, crossover, and mutation, evaluating each against a global fitness function (Koza, 1992). RL typically improves a single policy over time through trial-and-error interaction with the environment, using reward signals and temporal-difference methods for credit assignment (Sutton & Barto, 2018). While GP often produces interpretable, symbolic solutions, RL is generally more sample-efficient and better suited for problems requiring stepwise feedback (Brameier & Banzhaf, 2007).

Computational modelling techniques such as GEMS are beneficial for cognitive science, as they require formal specification of all aspects of a cognitive theory. All parts of a model/theory are defined, avoiding informal theorising that gives rise to the issues seen in attention research outlined above – attention is not an abstract placeholder for an effect, rather it is specified as a sequence of defined mechanisms. Using these techniques also reduces bias: researchers typically interpret results in light of their existing preconceptions regarding the topic. While researcher bias can certainly still feature in cognitive modelling (e.g. numerous decisions need to be made by the researchers regarding key aspects of modelling, each of which can introduce bias), it should be reduced relative to typical data interpretation, and can be reduced further by including more data, more operators and collaborating across research teams. These techniques further allow an efficient method to test different theories – for example, if an experiment that is used to investigate attentional effects can be explained by unrelated mechanisms, this suggests we cannot take for granted that our experiments are investigating what we think they are. Finally, it is an efficient way to generate novel ideas in a field with a wealth of data and cumbersome theories.

In the remainder of this paper, we use GEMS to generate models of attention in three experiments: one experiment using the central cueing paradigm (Arjona et al., 2016) and two experiments using the peripheral cueing paradigm (Langley et al., 2011; Lim et al., 2018). Table 1 provides information about the experimental and simulation details.

3. Method

3.1. GEMS Implementation

To evolve candidate models of human attention, we employed the GEMS system, a semi-automated method that uses the evolutionary principles of GP (for more details regarding GEMS, see Frias-Martinez and Gobet (2007), Lane et al. (2016)). Models are sequences of operators, each of which reflects cognitive operations and interacts with the GEMS architecture (see Table 2 for details of the operators included in the simulations described in this paper). Each model is run through the specific cueing experiment under consideration, with operators engaging with the cognitive architecture in specific ways. The (simple) cognitive architecture used in this paper is consistent across experiments and features core cognitive components. Namely: (i) a short-term-memory (STM) store with three slots, (ii) the location of attention, (iii) a 'current' buffer where currently attended items are stored, (iv) a salience map for salient areas of the visual field, (v) a preparatory response buffer, for storing the currently primed response, (vi) a response buffer to save the given response, (vii) a strength-association value, and (viii) a clock. Operators might have an effect on the cognitive architecture, and the state of the cognitive architecture during a trial might affect the operators within a model. For example, the respond-X operator sets the model response to X, and increases the model clock based on the value in the preparatory response buffer – if they do not match, then the prep-response-X operator is called, increasing the model clock further.

The strength-association value is updated each trial based on the validity of the trial, increasing following a valid cue, and decreasing following an invalid cue. The change of value is based upon the Rescorla-Wagner (RW) equation (Rescorla & Wagner, 1972) of associative learning, which determines the strength of the association between a signal (in this case, the cue) and a stimulus (the target). This value remained close to 0.5 for the peripheral experiments (Langley et al., 2011; Lim et al., 2018), as cue validity was 50 %.

Table 1
Details of GEMS modelled cueing experiments.

	Arjona et al. (2016)	Langley et al. (2011)	Lim et al. (2018)
Published Experiment Details			
Number of trials	300	80	360
Participant age range	18–35	18–28	18–28
Type of cue	Central arrow	Peripheral	Peripheral
Cue validity	50 %, 68 %, 86 %	50 %	50 %
Cue modality	Visual	Visual	Visual
Target modality	Auditory	Visual	Visual
Response type	Button press	Button press	Eye movement
Cue phase (ms)	300	50	100
Interstimulus interval (ms)	370	50, 250, 550, 950	150, 300, 450, 600, 750, 900
Response window (seconds)	2	6	3
Simulation Details			
Fitness weights (accuracy, RT, model size)	0.45, 0.5, 0.05	0.3, 0.65, 0.05	0.3, 0.65, 0.05
Best model fitness	0.03	0.03	0.03
RMSE	11.68	8.45	12.82
R^2	0.70	0.90	0.82

Table 2
Overview of the operators used by GEMS in this paper's simulations. Each operator type had a set time (in milliseconds, ms) gleaned from the scientific literature, as follows: input (100 ms), output (70 ms), cognitive (70 ms), STM (50 ms), syntax (0 ms). The subscript 'v' indicates that the timing is variable and can be altered via other factors.

Name	Function	Type
attend	Puts the item at the attention location into model 'current'	input
move-att-X	Shift 'attention' to a location in the visual display ($X \in \{centre, left, right, clockwise, counterclockwise\}$)	cognitive
attn-capture	if there is a stimulus, move attention to its location	cognitive
current-X-p	Predicate for stimulus type of model 'current' ($X \in \{target, stimulus\}$)	cognitive
prep-response-X	Sets model 'prep-response' to X, advances model clock based on saliency-map ($X \in \{left, right, centre\}$)	cognitive$_v$
respond-X	Sets model 'response' to X ($X \in \{left, right, centre,$ 'current'-item-location$\}$) executes prep-response-X if X doesn't match 'prep-response'	output
rehearsal-N	Updates item-time in STM item N ($N \in \{1, 2, 3\}$) to current model 'clock'	stm
retrieve-N	Sets model 'current' to STM item N ($N \in \{1, 2, 3\}$)	stm
retrieve-X	Sets model 'current' to STM item matching type X ($X \in \{target, stimulus\}$)	cognitive
nil	Sets model 'current' to nil	cognitive
put-stm	Pushes value in model 'current' and model 'clock' to STM slot 1	cognitive
RW-cue-strength	Given a cue, predict target location and prepare response	cognitive$_v$
RW-cue-percept	Give a cue and target, predict target location and prepare response	cognitive$_v$
if-strength-assoc	If model 'strength-assoc' over threshold 0.7, respond in line with cue	output
prev-val	If the previous trial was valid, respond in line with cue	output
dotimes-N	Repeats a given expression ($N \in \{2, 3, 4\}$)	syntax
if	Executes condition, executes one of two expressions depending on the condition	syntax
prog-N	Sequence of expressions ($N \in \{2, 3, 4\}$)	syntax
wait-N	Advances model clock by the specified number N of milliseconds ($N \in \{25, 50,$ $100, 200, 1000, 1500,$ 0.5-trial-length, 0.25-trial-length, 0.1-trial-length$\}$)	syntax
while-N	Repeats an expression for a set time in ms ($N \in \{100, 200\}$)	syntax

For this implementation of the cueing experiment, STM had three slots, in order to keep the model structure simple. However, GEMS is a flexible system, and different numbers of slots will be explored in future work.

Models are constrained so that only one response can be given – once the response buffer has been filled, the program will end. The structure of the architecture is invariant across experiments and across GP runs: GP is not changing the architecture, rather it is changing the models that are implemented within the architecture. The reaction time and accuracy of each model is stored and averaged for each experimental condition (ISI for the peripheral experiments, block validity for the arrow experiment). These values are then compared against the averaged reaction times and accuracy of the published datasets.

A multi-objective fitness function, within the range 0–1, indicates how well the model data fits the published data, where a smaller value indicates a better fit. In particular, the difference between the model and published data (reaction time and accuracy) for each experimental condition is scaled to the range 0–1. (RT was scaled to 0–1 using a half-sigmoid function.) These comparisons are then weighted and summed to generate a single fitness value. The computation of fitness is phased during evolution (see Lane et al. (2022)) so that models are first optimised for accuracy, then accuracy and reaction time, followed by accuracy, reaction time and model size (preferring smaller models); in the final phase each parameter is weighted (see Table 1 for fitness weights). Each experimental condition was equally weighted.

The models from the final generation with the best fitness values were then processed, replacing operators that were only advancing the

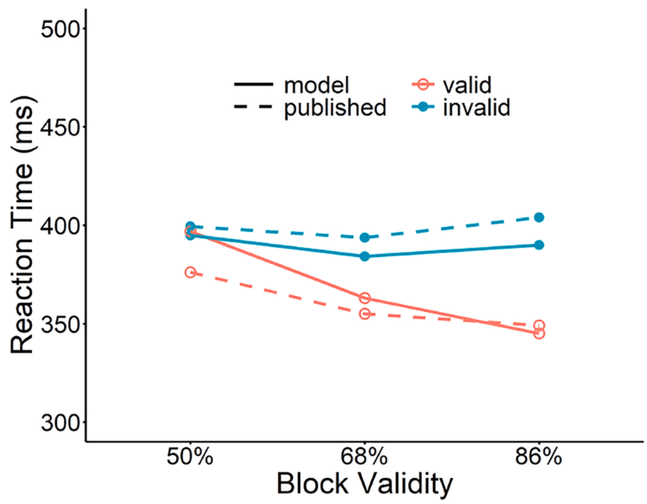

Fig. 2. Reaction times for the published behavioural data and model output for each experimental condition of Arjona et al. (2016).

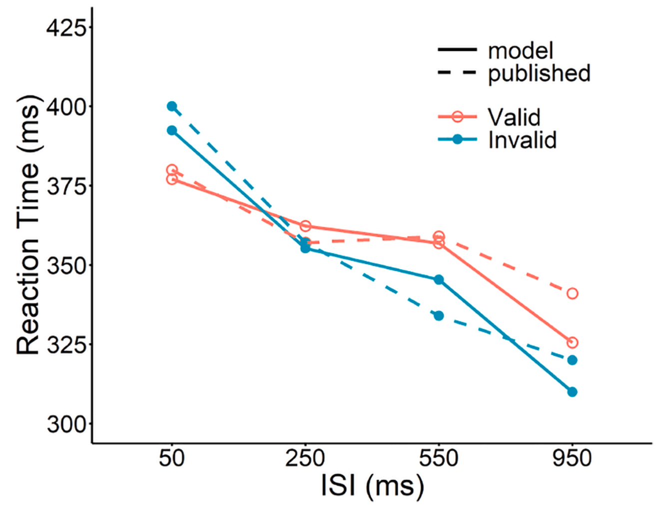

Fig. 3. Reaction times for the published behavioural data and model output for each experimental condition of Langley et al. (2011).

clock with `wait` operators (see Lane and Gobet (2024) for more details of the post-processing techniques). This allowed the removal of any duplicate models from the set, resulting in a more interpretable set of models.

3.2. Published experimental data

Three experiments were modelled using the GEMS method: one using the centrally presented cueing paradigm, and two with peripheral cues. Across these experiments we explored the effect of cue type (central vs. peripheral), response method (button press vs. eye movement), a wide range of ISIs, different durations for both cue and target presentations, and the effect of cue informativeness. For all experiments, a population of 3000 models were evolved over 500 generations.

3.2.1. Central cueing paradigm

In the experiment carried out by Arjona et al. (2016), an arrow cue was presented centrally and an auditory target stimulus was presented to the left or the right ear. The duration of each phase can be found in Table 1. To record their response, participants pressed a button indicating the location of the target.

Three cue-validity conditions allowed analysis of learning across a block of trials: in one condition the cue was valid 50 % of the time (uninformative), another had cues that were valid for 68 % of the trials, and another was 86 % valid (highly informative cue). Participants were not informed of the condition they were in. As can be seen in Fig. 2, there is a cueing effect for all block validity conditions; participants responded more quickly following a valid cue compared to an invalid cue. Further, this cueing effect increased with increasing block validity, moving from only a very slight effect for nonpredictive cues (the 50 % condition) to a slightly greater effect when the cue was valid for 86 % of trials in a block.

In order to model learning, operators based on the Rescorla-Wagner equation, which determines the associative strength of a cue and a target, were included. Specifically, either using the cue alone, or both the cue and target together, these operators predicted where the target was likely to appear.

3.2.2. Peripheral cueing with button response

To simulate IOR and peripheral cueing, we used the experiment carried out by Langley et al. (2011). The duration for each phase of the experiment can be found in Table 1. Participants provided a response by pressing the appropriate button indicating the location of the target. While the published paper included a number of other conditions (three different age bands of participants, removal of the delay phase), we only included data for the youngest group and with a delay phase, to align with the other experiments modelled.

This experiment manipulated the duration of the interval between the cue and the target stimuli (the ISI); as can be seen in Fig. 3, following an initial facilitatory effect at shorter ISIs, the pattern reduces and is reversed for longer ISIs. With increasing ISI, RTs reduce – participants are quicker to respond when the delay phase is longer regardless of the validity of the trial.

3.2.3. Peripheral cueing with eye movement response

A second peripheral cueing experiment (Lim et al., 2018) was used to generate models of IOR with a broader range of ISIs (see Table 1 for details) and with a different method of response: eye movements towards a target were recorded rather than a button press, resulting in shorter RTs compared to both experiments described above. The timing for the response operator was reduced to 0 ms for the experiment of Lim et al. (2018), as the hand motor response was absent.

The pattern of results for this experiment (which can be seen in Fig. 4) differs slightly from the Langley et al. (2011) peripheral experiment as there is no evidence of a facilitation effect at short ISIs. In particular, the shortest ISI of this experiment was longer than those of Langley et al. (2011), and so this initial facilitation was not demonstrated. Rather, the IOR effect (quicker responses for invalid cues) increases between the two shortest ISIs, and then stays mostly constant. RTs for both valid and invalid trials are reduced as ISI increases, as in Langley et al. (2011).

4. Simulations

4.1. Central cueing and learning models

For the Arjona et al. (2016) experiment, initial models were generated to capture the 86 % valid block effect, as this condition had the greatest effect of cue validity. The best model for this condition was then used to seed 10 % of the initial 3000 models generated for simulating the results of the entire experiment (with all three block conditions), with the remaining 90 % of models randomly generated.

Post-processing of the final generation of models with the best fitness values resulted in one final model with a fitness of 0.02 (0 indicates a perfect fitness), which can be seen in Fig. 5. In order to account for randomness in some operators, this final model was rerun 1000 times; the averaged RTs for each condition can be seen in Fig. 2.

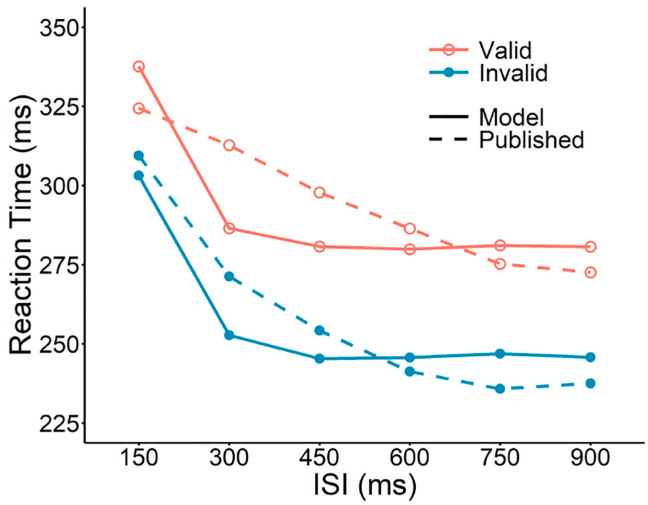

Fig. 4. Reaction times for the published behavioural data and model output for each experimental condition of Lim et al. (2018).

This model particularly utilised the `RW-cue-strength` operator, which predicts the target location and prepares a response based on the cue. Initially, the model waits and then attends the arrow cue, then uses `RW-cue-strength` to prepare a response given the cue and the strength association of the block. The model then moves attention right, attends again and responds with the currently attended item (if there is anything currently attended). This attend and respond-with-attended cycle is then repeated if the trial is continuing (i.e. there is no target stimulus on the right), until the model finally responds left. This is an interesting strategy for the cueing task – it is essentially localising attention on one side, and if a target stimulus never appears, after a certain time it responds in line with the unattended side. This model follows the pattern of results demonstrated in Arjona et al. (2016), with an increasing effect of cue validity as the cue becomes more informative.

4.2. Peripheral cueing with button response models

Initial models generated for the Langley et al. (2011) experiment captured the IOR effect; however, they did not demonstrate the facilitation effect at shorter ISI's. To address this issue, models for just the 50 ms and 250 ms ISI conditions were evolved. The best fitting model for these conditions (with a fitness of 0.01) then comprised 10 % of the initial 3000 models generated for the full dataset (with all four ISI conditions).

Post-processing of the models from the final generation, followed by removal of duplicate models from the set, resulted in one final model with a fitness of 0.03 (seen in Fig. 6).

The best model for Langley et al. (2011) utilised the `dotimes` operators to cycle through sequences of searching for a stimulus in the display (attention capture) and responding in line with any stimulus that has been found. If no stimulus is found, the cycle repeats, until a response has been given. If no target is ever detected, the model will end by responding ‘right’.

The model, like the human data, demonstrated faster reaction times with increasing ISIs (see Fig. 3). The facilitation effect seen in the human data for the shortest interval of 50 ms and the minimal difference between valid and invalid trials for the 250 ms ISI was also captured by the model. Further, the inhibition of return effect was evident – that is, an invalid cue resulted in faster reaction times than a valid cue, for the 550 ms and 950 ms intervals. As such, this model captured the main features of the original human data. It is interesting to note that this pattern of results was found even though the cue, which is central to this paradigm, was ignored.

4.3. Peripheral cueing with eye movement response

Following post-processing, the best model for the Lim et al. (2018) experiment had a fitness of 0.03, and can be seen in Fig. 7.

This model first attempts to store the cue, by moving attention to the left and then attending. The model then prepares a response to the left. Through using the RW-cue-strength operator, if the cue was attended (i.e., it was presented to the left), Rescola-Wagner learning operations based on the model strength association are used to predict the likely target location, and a response is prepared. The model then utilises the dotimes operators to cycle through attention capture, attending and responding with anything that is attended.

This model is consistent with the published data, with a clear IOR effect (see Fig. 4) – valid trials resulted in slower reaction times than invalid trials. Further, the IOR effect increased from the shortest ISIs, and remained approximately consistent following the 300ms ISI. While the model captures these trends, it results in a more dramatic effect of ISI than the more gradual slowing of RT seen in the published data.

4.4. Model comparison

For each of the cueing experiments described, the same architecture was used and the GEMS system had access to the same operators. This demonstrates that the system is capable of producing models with a good fit for different kinds of tasks. The models themselves have some similarities across the experimental paradigms. Namely, a cycle of attending and responding in line with what is attended was prevalent across all models. The central arrow model focused on one side of the environment, with a final response to the opposite side if no target stimulus is detected, whereas the peripheral cuing models utilised the attention capture operator, more in line with the exogenous type of orienting. While the models simulating the experiments carried out by Arjona et al. (2016) and Lim et al. (2018) utilised the cue, the model simulating the experiment by Langley et al. (2011) did not – this is likely due to the very short duration that the cue was presented for in the experiment. Operator timings could be altered in the future to allow perception of the cue; however, interestingly the model still demonstrates the facilitation and IOR pattern of the published results. It could be the case, therefore, that other elements of the experimental design, or simpler cognitive mechanisms are driving the effects seen in cueing experiments. This is one of the benefits of using a system such as GEMS – we can explore whether experimental paradigms are actually investigating what they set out to.

5. Discussion

Issues within psychology in general and attention research in particular, such as inconsistency in definition and informal, verbal theories, require new approaches and a critical evaluation of how theories are formed. Computational modelling, while not a new technique, is increasingly accessible to psychologists, and could benefit the field. Due to the nature of modelling, theories must be well specified, avoiding the kind of informal theories often used in psychology. Advances in AI can further be applied to psychological research. Here, using GEMS, a new method of theory generation, we were able to capture a number of the key behavioural findings for the attentional cueing paradigm. In particular, we generated models for both central and peripheral cueing experiments, encapsulating patterns of facilitation and inhibition.

The models generated by GEMS for the cueing paradigm were able to capture the pattern of results seen for human participants. Specifically, facilitation effects were demonstrated with short ISI peripheral cueing (Langley et al., 2011) and with centrally presented arrow cues (Arjona et al., 2016). Further, the increased facilitation effect when a central arrow cue is more likely to predict the location of a target was demonstrated (Arjona et al., 2016). Finally, the IOR effect, with slower RTs given a valid cue in a peripheral cueing experiment at longer ISIs was demonstrated in both Langley et al. (2011) and Lim et al. (2018).

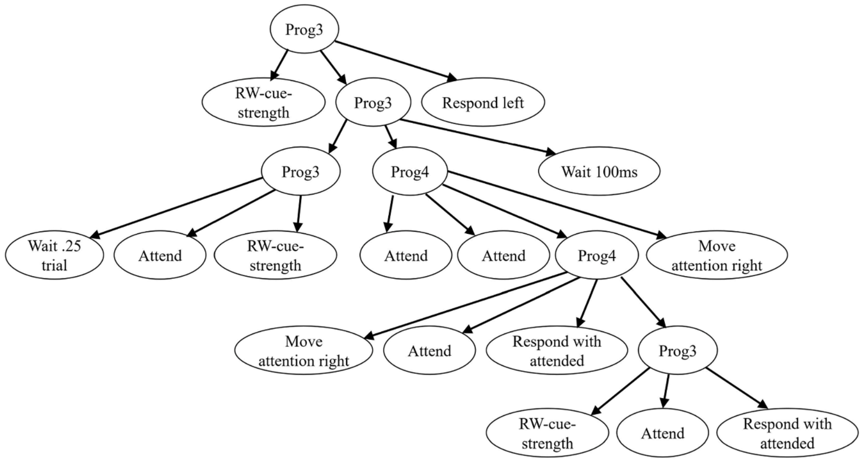

Fig. 5. Graphical representation of the best-fitting model generated by GEMS for Arjona et al. (2016). Fitness = 0.03.

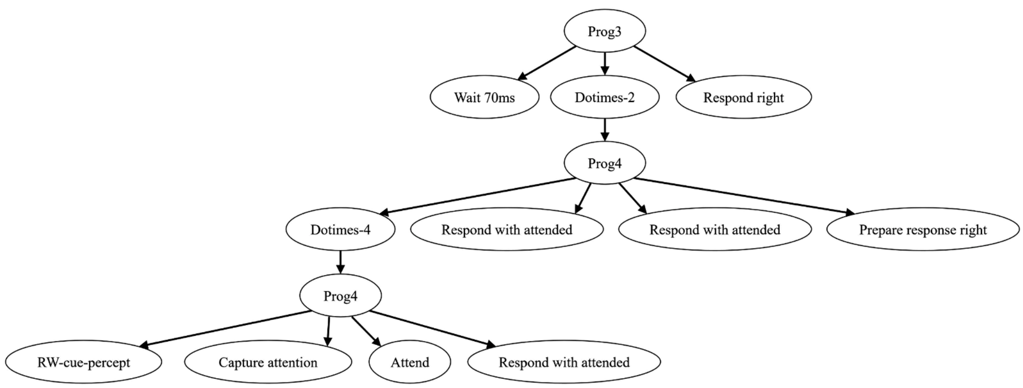

Fig. 6. Representation of the best-fitting model generated by GEMS for Langley et al. (2011). Fitness = 0.03.

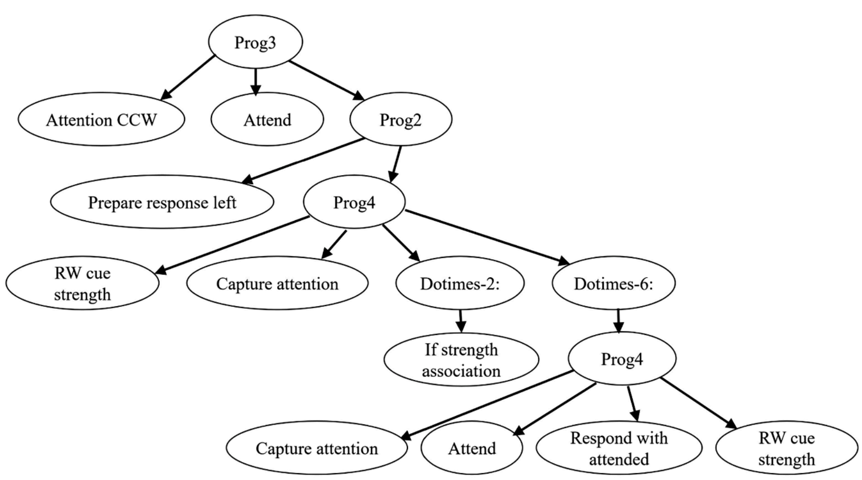

Fig. 7. Representation of the best-fitting model generated by GEMS for Lim et al. (2018). Fitness = 0.03.

This suggests that GEMS is a capable method of producing well-defined models and strategies for cognitive tasks.

Although models produced for Arjona et al. (2016) and Lim et al. (2018) datasets utilised the cue, the model for Langley et al. (2011) did not, while still demonstrating the facilitation and IOR pattern. This brings into question what behaviour is happening in these experimental paradigms, and whether interpretations relying on attention and cueing should be more critically considered. On the one hand, the cue may be ignored due to decisions about the timings of operators. For this dataset, the cue was presented very briefly (only 50 ms), while most of the operators take more time than this. These timings were taken from the literature; however, future work will determine whether these timings should be adjusted. On the other hand, that GEMS was able to produce models that fit the human data despite not acknowledging the cue suggests that the findings of such experiments could be due to the exploitation of different features of the cognitive system. More focused experimentation is necessary to further explore these considerations.

GEMS assists in theory development in a multitude of ways. For the current paper, the best fitting model for both Arjona et al. (2016) and Langley et al. (2011) featured an unexpected strategy (focusing attention on just one side of the environment, and ignoring the cue entirely), which casts doubt on the typical interpretations of the cueing paradigm. GEMS's capacity to generate surprising strategies, as noted in earlier work, appears to stem in part from its ability to exploit idiosyncratic features of the experimental settings. These "low-level" strategies illustrate the value of genetic programming for theory development while also highlighting potential limitations of researchers in exploring the hypothesis space. An important open question is whether human participants employ comparable strategies, a topic that warrants further investigation.

An additional benefit is that, for generating new theories, the architecture and the operators used by the GEMS system have to be well specified, which results in clear definitions of key concepts. For example, attention is operationalised by operators that have direct effects that can be examined, adjusted and explored. For the models generated in the current paper, attention was treated as a point that can be moved around a display in a number of ways, where the area receiving attention is salient. However, it is possible to define attention in many different ways, as well as compare these theories against one another. By providing the system with different options, we can explore differences in fitness values. All attention-related operators could also be removed, to see if the effects being measured by tasks designed to isolate attention can be explained by other mechanisms.

As shown in the current paper, the GEMS system can be applied to multiple datasets. In a previous example, Bartlett et al. (2023c) applied the methodology to two datasets for the delayed-match-to-sample task, to determine the importance of often taken-for-granted factors in experimental methods (such as the duration of stimulus presentation and the ISI). With the ever-increasing amount of data being collected, it is useful to integrate findings to get a more holistic view of what mechanisms might be underlying similar and dissimilar behaviours. By using multiple experiments, we have been able to identify similarities and differences between models for different experimental conditions, allowing a better understanding of the mechanisms of the cueing experiment.

A key benefit of GEMS is that the potential for bias is reduced. In particular, researchers in a given field might have particular assumptions, which would affect how data is interpreted, as well as what research is conducted and how it is designed. While GEMS will still have bias in terms of how the architecture is structured, which operators are included and how they are operationalised, these can be reduced by feeding in different operators, different definitions and investigating different theories. Further, the models that GEMS produces are novel and can be unexpected. Such models can then suggest directions for experimental research that may not have been considered previously. One of the criticisms of psychology in general and attention research in particular is that those working in a particular domain might not appreciate the impact of other mechanisms on behaviour. For example, a task assumed to measure short-term memory might also engage mechanisms relating to decision making. Including operators from many domains avoids this, and encourages collaboration between researchers.

A few limitations of the approach might be noted. First, we had to experiment with the three weights in the fitness function in order to get good models; hence, the weights differ between the three experiments. Second, the modeller had to perform some preliminary search for helping GEMS find suitable models. For example, for the first two experiments, a preliminary run had to be carried out with a subset of the conditions – the ones with the largest cueing effect – in order to find models that could partly seed the simulations for all the conditions. Third, because the smallest non-zero time parameter was 50 ms, it was not possible to achieve a perfect fit for the RT component of the fitness function. Optimising the parameters alongside the models would resolve this limitation. Fourth, a related issue is that, despite the large search space, all three experiments converged on a single best-fitting model (after post-processing and de-duplication). The relatively small number of operators, combined with the strong constraints imposed by the timings, appears to have drastically restricted the solution space. Once again, optimising the timing parameters – along with expanding the set of operators – may help to address this problem.

Searching for a model that can explain data from multiple datasets is a critical future application of GEMS. As timings often differ across experiments (as in the current paper), timing parameters will need to be optimised in future work to achieve this. Additional methods for analysing the models generated by GEMS will also be developed in future applications. For example, analysing how each specific operator influences the behaviour of the model would allow better understanding of how each operator relates to higher level attentional processes. This could potentially be achieved by removing each operator in the model and more thoroughly analysing the effect of its removal on model outputs. This would also help us to determine which operators are most critical for the experiment. While the models generated by GEMS have been displayed in tree diagrams, clearer and more interpretable methods of displaying models will be explored in the future. Further, the integration of GEMS with more complex and well established architectures is an important future direction. This would allow a powerful and nuanced method for modelling the complex field of attention, building upon the critical work undergone by those refining detailed cognitive architectures.

The flexibility of the GEMS system and the benefits outlined above suggest that it could be a useful tool for theory development within psychology, addressing the lack of clearly defined and well specified theories.

CRediT authorship contribution statement

Laura K. Bartlett: Writing – review & editing, Writing – original draft, Visualization, Formal analysis; **Noman Javed:** Writing – review & editing; **Dmitry Bennett:** Writing – review & editing; **Peter C.R. Lane:** Writing – review & editing, Software, Resources, Methodology; **Fernand Gobet:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Funding from the European Research Council (ERC-ADG-835002—GEMS) is gratefully acknowledged.

References

Anderson, B. (2011). There is no such thing as attention. *Frontiers in Psychology*, *2*, 1–8. https://doi.org/10.3389/fpsyg.2011.00246

Anderson, B. (2021). Stop paying attention to "attention". *Wiley Interdisciplinary Reviews: Cognitive Science*, *14*, 1–6. https://doi.org/10.1002/wcs.1574

Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., & Qin, Y., et al. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060. https://doi.org/10.1037/0033-295X.111.4.1036

Arjona, A., Escudero, M., & Gómez, C.M., et al. (2016). Cue validity probability influences neural processing of targets. *Biological Psychology*, *119*, 171–183. https://doi.org/10.1016/j.biopsycho.2016.07.001

Bartlett, L.K., Pirrone, A., Javed, N., & Gobet, F. (2023a). Computational scientific discovery in psychology. *Perspectives on Psychological Science*, *18*, 178–189. https://doi.org/10.1177/174569162210918

Bartlett, L.K., Pirrone, A., Javed, N., Lane, P., & Gobet, F. (2023b). Generating candidate cognitive models for the posner cueing task. In *45th European conference on visual perception*.

Bartlett, L.K., Pirrone, A., Javed, N., Lane, P.C.R., & Gobet, F. (2023c). Genetic programming for developing simple cognitive models. In M. Goldwater, F.K. Anggoro, B. K. Hayes, & D.C. Ong (Eds.), *Proceedings of the 45th annual conference of the cognitive science society* (pp. 2833–2839).

Brameier, M.F., & Banzhaf, W. (2007). Linear genetic programming. Springer.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012

Chica, A.B., Martín-Arévalo, E., Botta, F., & Lupiáñez, J. (2014). The spatial orienting paradigm: How to design and interpret spatial attention experiments. *Neuroscience and Biobehavioral Reviews*, *40*, 35–51. https://doi.org/10.1016/j.neubiorev.2014.01.002

Dukewich, K.R., & Klein, R.M. (2015). Inhibition of return: A phenomenon in search of a definition and a theoretical framework. *Attention, Perception, and Psychophysics*, *77*, 1647–1658. https://doi.org/10.3758/s13414-015-0835-3

Eriksen, B.A., & Eriksen, C.W. (1974). Effects of noise letters upon the identification of a target letter in a non search task. *Perception and Psychophysics*, *16*, 143–149.

Eriksen, C.W., & St. James, J.D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, *40*(4), 225–240. https://doi.org/10.3758/BF03211502

Fan, J., McCandliss, B.D., Sommer, T., Raz, A., & Posner, M.I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, *14*(3), 340–347. https://doi.org/10.1162/089892902317361886

Frias-Martinez, E., & Gobet, F. (2007). Automatic generation of cognitive theories using genetic programming. *Minds and Machines*, *17*(3), 287–309. https://doi.org/10.1007/s11023-007-9070-6

Hilchey, M.D., Klein, R.M., & Satel, J. (2014). Returning to "inhibition of return" by dissociating long-term oculomotor IOR from short-term sensory adaptation and other nonoculomotor "inhibitory" cueing effects. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 1603–1616. https://doi.org/10.1037/a0036859

Hommel, B., Chapman, C.S., Cisek, P., Neyedli, H.F., Song, J.H., & Welsh, T.N. (2019). No one knows what attention is. *Attention, Perception, and Psychophysics*, *81*, 2288–2303. https://doi.org/10.3758/s13414-019-01846-w

Hussain, F., & Wood, S. (2009). Modelling the efficiencies and interactions of attentional networks. In L. Paletta, & J.K. Tsotsos (Eds.), *Attention in cognitive systems. WAPCV 2008. lecture notes in computer science* (pp. 139–152). (*vol. 5395*). https://doi.org/10.1007/978-3-642-00582-4_11

King, R.D., Whelan, K.E., Jones, F.M., Reiser, P.G.K., Bryant, C.H., Muggleton, S.H., Kell, D.B., & Oliver, S.G. (2004). Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature*, *427*, 247–252. https://doi.org/10.1038/nature02236

Klein, R.M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, *4*, 138–147. https://doi.org/10.1016/S1364-6613(00)01452-2

Lane, P.C.R., Bartlett, L.K., Javed, N., Pirrone, A., & Gobet, F. (2022). Evolving understandable cognitive models. In *International conference on cognitive modeling*.

Lane, P.C.R., & Gobet, F. (2024). Tutorial: Computational scientific discovery in cognitive science. In *Proceedings of the IEEE 6th international conference on cognitive machine intelligence*.

Lane, P.C.R., Sozou, P.D., Gobet, F., & Addis, M., et al. (2016). Analysing psychological data by evolving computational models. In A. Wilhelm, & H. Kestler (Eds.), *Analysis of large and complex data. studies in classification, data analysis, and knowledge organization, proceedings of ECDA 2014* (pp. 587–597). Springer. https://doi.org/10.1007/978-3-319-25226-1

Langley, L.K., Friesen, C.K., Saville, A.L., & Ciernia, A.T. (2011). Timing of reflexive visuospatial orienting in young, young-old, and old-old adults. *Attention, Perception, & Psychophysics*, *73*, 1546–1561. https://doi.org/10.3758/s13414-011-0108-8

Lim, A., Eng, V., Janssen, S.M.J., & Satel, J. (2018). Sensory adaptation and inhibition of return: dissociating multiple inhibitory cueing effects. *Experimental Brain Research*, *236*, 1369–1382. https://doi.org/10.1007/s00221-018-5225-3

Lupianez, J., Klein, R.M., & Bartolomeo, P. (2006). Inhibition of return: Twenty years after. *Cognitive Neuropsychology*, *23*, 1003–1014. https://doi.org/10.1080/02643290600588095

Lupianez, J., Milán, E.G., Tornay, F.J., Madrid, E., & Tudela, P. (1997). Does IOR occur in discrimination tasks? Yes, it does, but later. *Perception & Psychophysics*, *59*, 1241–1254. https://doi.org/10.3758/BF03214211

Mozer, M.C. (1991). The perception of multiple objects: A connectionist approach. The MIT Press.

Mozer, M.C., & Sitton, M. (1998). Computational modeling of spatial attention. In H. Pashler (Ed.), *Attention* (pp. 341–393). London: Psychology Press.

Musslick, S., Bartlett, L.K., Chandramouli, S.H., Dubova, M., Gobet, F., Griffiths, T.L., Hullman, J., King, R.D., Kutz, J.N., Lucas, C.G., Mahesh, S., Pestilli, F., Sloman, S., & Holmes, W. (2025). Automating the practice of science: Opportunities, challenges, and implications. *Proceedings of the National Academy of Sciences*, *122*(5), e240123812.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). https://doi.org/10.1126/science.aac4716

Peterson, J.C., Bourgin, D.D., Agrawal, M., Reichman, D., & Griffiths, T.L. (2021). Using large-scale experiments and machine learning to discover theories of human decision-making. *Science*, *372*, 1209–1214. https://doi.org/10.1126/science.abe2629

Posner, M.I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25. https://doi.org/10.1080/00335558008248231

Posner, M.I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma, & D. Bonwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 531–556). Erlbaum.

Posner, M.I., & Petersen, S.E. (1990). The attention systems of the human brain. *Annual Review of Neuroscience*, *13*, 25–42.

Posner, M.I., Rafal, R.D., Choate, L.S., & Vaughan, J. (1985). Inhibition of return: Neural basis and function. *Cognitive Neuropsychology*, *2*, 211–228. https://doi.org/10.1080/02643298508252866

Posner, M.I., Snyder, R.R., & Davidson, B.J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160–174. https://doi.org/10.1037/0096-3445.109.2.160

Remington, R., & Pierce, L. (1984). Moving attention: Evidence for time-invariant shifts of visual selective attention. *Perception & Psychophysics*, *35*(4), 393–399. https://doi.org/10.3758/BF03206344

Rescorla, R.A., & Wagner, A.R. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A.H. Black, & W.F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton Century Crofts.

Ristic, J., Friesen, C.K., & Kingstone, A. (2002). Are eyes special? It depends on how you look at it. *Psychonomic Bulletin & Review*, *9*, 507–513.

Rosenholtz, R. (2018). Rethinking capacity limits in visual processing: peripheral vision, attention, and decision limits. *Journal of Vision*, *18*. https://doi.org/10.1167/18.10.1185

Samuel, A.G., & Kat, D. (2003). Inhibition of return: A graphical meta-analysis of its time course and an empirical test of its temporal and spatial properties. *Psychonomic Bulletin & Review*, *10*, 897–906. https://doi.org/10.3758/BF03196550

Satel, J., Wilson, N.R., & Klein, R.M. (2019). What neuroscientific studies tell us about inhibition of return. *Vision*, *3*, 1–13. https://doi.org/10.3390/vision3040058

Sutton, R.S., & Barto, A.G. (2018). Reinforcement learning. The MIT Press.

Taylor, H. (2023). Attention as a patchwork concept. *European Journal for Philosophy of Science*, *13*. https://doi.org/10.1007/s13194-023-00538-5

Wang, H., Fan, J., & Johnson, T.R. (2004). A symbolic model of human attentional networks. *Cognitive Systems Research*, *5*(2), 119–134. https://doi.org/10.1016/j.cogsys.2004.01.001