Atenção seletiva e informação de alto nível: modelos de seleção da informação em cenas naturais

Frederico Miranda Rodrigues Pinheiro – Universidade Federal de Uberlândia, Uberlândia, Minas Gerais, Brasil
Joaquim Carlos Rossini – Universidade Federal de Uberlândia, Uberlândia, Minas Gerais, Brasil

Resumo
De modo geral, as investigações acerca da atenção seletiva apresentaram uma grande ênfase no estudo do processamento da informação visual de baixo nível e médio nível, fazendo uso, para tanto, de estímulos visuais relativamente simples (letras, números e formas geométricas). No entanto, tem crescido o número de estudos sobre a influência da informação de alto nível no processamento de cenas mais complexas e contextualizadas. A presente revisão tem como objetivo apresentar estudos relevantes sobre este tema e centra a sua discussão nos principais aspectos divergentes entre os modelos discretos do processamento atentivo, que preveem que a informação de alto-nível é processada em momentos tardios, e os modelos de processamento inicial, que preveem uma influência da informação de alto-nível em momentos precoces do processo de seleção da informação visual.
Palavras-chave: Atenção seletiva; Busca visual; Informação de alto nível.

Selective Attention and high-level information: model of information selection in natural scenes

Abstract
The researches about selective attention showed great emphasis on the study of low-level and mid-level processing of visual information, using relatively simple visual stimulus (letters, numbers and geometric shapes). However, there are an increasing number of studies about the influence of high-level information on the selection with complex scenes in natural contexts. The aim to present article is review the relevant studies about this area and highlight the differences between the discrete process model, which predicts that high-level information is processed in the late stages, and the early model that predicts an influence of high-level information on early visual selection.
Keywords: Selective attention; Visual search; High-level information.

Atención selectiva e informaciones de alto-nivel: modelos de selección de la información en escenas naturales

Resumen
En general, las investigaciones respecto a la atención selectiva presentaron gran énfasis en el estudio del procesamiento de la información visual de bajo nivel y medio nivel, haciendo uso, para tanto, de estímulos visuales relativamente simples (letras, números y formas geométricas). Sin embargo, ha crecido el número de estudios sobre la influencia de la información de alto nivel en el procesamiento de escenas más complejas y contextualizadas. La presente revisión tiene como objetivo presentar estudios relevantes sobre este tema y centrar su discusión en los principales aspectos divergentes entre los modelos discretos del procesamiento atentivo, que predicen que la información de alto nivel es procesada en momentos tardíos, y los modelos de procesamiento inicial, que predicen una influencia de la información de alto nivel en momentos precoces del proceso de selección de la información visual.
Palabras-clave: Atención selectiva; Busca visual; Información de alto nivel.

Enquanto dirigimos nosso carro pelo centro de uma grande cidade, somos frequentemente distraídos por uma grande quantidade de estímulos, como outros carros, *outdoors* ou pedestres. Nessa atividade, selecionamos estímulos relevantes e descartamos estímulos irrelevantes contidos em um ambiente saturado de informação para chegarmos com segurança ao nosso destino. Esse processo de seleção da informação é o que caracteriza a atenção seletiva. O estudo da atenção seletiva é comumente realizado por meio de tarefas de busca visual em que os participantes procuram ativamente por um determinado estímulo-alvo apresentado em meio a estímulos distratores. Durante esse processo, o sucesso em localizar o alvo é decorrente de três fontes de informações: a informação de baixo nível, a informação de médio nível e a informação de alto nível (Henderson & Hollingworth, 1999). A informação de baixo nível é caracterizada pelas características físicas simples dos estímulos, tais como cor, forma, luminância e movimento. A informação de médio nível é o resultado da integração dessas características físicas simples, possibilitando a percepção dos objetos, mas ainda sem significado contextual. Por último, a informação de alto nível refere-se à informação semântica dos estímulos. Esse processamento é efetuado pelos recursos da memória de curto e longo prazo e possibilita o reconhecimento contextual de cenas complexas (Henderson & Hollingworth, 1999).

Uma teoria bastante influente sobre o processo de seleção e análise da informação visual foi proposta no inicio da década de 1980 e ficou conhecida como Teoria da Integração das Características (TIC) (Treisman & Gelade, 1980). A TIC propõe que a busca visual ocorre em, pelo menos, dois estágios sequenciais e distintos, sendo um pré-atentivo e o outro atentivo. No estágio pré-atentivo, o processamento da informação ocorre de forma paralela por todo campo visual, possibilitando a rápida identificação de um alvo definido por uma característica única. No entanto, quando o estímulo-alvo compartilha várias características com os estímulos distratores, é

necessário o engajamento de recursos atentivos, o que caracteriza o estágio atentivo. Esses recursos integram as características únicas em objetos, possibilitando a identificação do alvo. Assim, os objetos complexos de uma cena visual seriam integrados um após o outro, em um processo serial e aleatório, até a identificação do alvo.

Paralelamente ao desenvolvimento da TIC, outras teorias foram elaboradas com o intuito de explicar aspectos do processo de seleção até então não abordados. Um bom exemplo nesse sentido é a Teoria da Similaridade proposta por Duncan e Humphrey (1989). Essa abordagem propõe que a dificuldade em localizar um determinado alvo não é expressa apenas em função do número de estímulos a serem integrados, mas sim em virtude de uma quantidade maior de fatores, como a similaridade entre os distratores e entre os distratores e o alvo. Na mesma época, Wolfe, Cave e Franzel (1989) propuseram outra teoria bastante influente, conhecida como Teoria da Busca Guiada (Wolfe, 2007; Wolfe, Cave & Franzel, 1989). Esse enfoque propõe que o processo de seleção pré-atentivo da informação é responsável por guiar os recursos atentivos na localização do alvo, baseando-se nas características físicas dos estímulos para reduzir a quantidade de prováveis alvos e, assim, aumentar a eficiência da busca.

Apesar das contribuições importantes de cada uma dessas teorias, todas abordam quase exclusivamente o processamento da informação de baixo nível e médio nível, e não avançam na investigação de como os estímulos reais e a informação de alto nível pode influenciar o processo de identificação de um estímulo-alvo. Na tentativa de abordar esse aspecto do processamento, algumas pesquisas, ainda na década de 1970, buscaram enfatizar a relevância do processamento de alto nível na percepção da informação (Biederman, 1972; Biederman, Glass & Stacy, 1973) fazendo uso de imagens extraídas do mundo real e que são, na maioria das vezes, processadas rapidamente.

Uma característica fundamental a ser considerada é que a memória dos ambientes é estruturada com base em modelos, que por sua vez possibilitam previsões probabilísticas das configurações esquemáticas gerais da cena (Biederman, 1972; Friedman, 1979). Esse modelo seria acessado no início da visualização e permitiria a ativação das representações de relações entre os objetos, facilitando, assim, o processamento da informação. Uma vez que os modelos clássicos da atenção preveem que a informação de alto nível é acessada tardiamente, duas questões surgem: 1) qual é a influência do contexto na seleção rápida da informação, e 2) como o acesso à informação dessas relações contextuais de alto nível pode ser suficientemente rápido para facilitar a percepção de um objeto em cenas naturais?

Biederman (1972) demonstrou a influência dos esquemas na percepção mediante um experimento em que os sujeitos observavam fotos de cenas reais em preto e branco em projeções de *slides*. Uma imagem de setas foi usada para indicar a área-alvo. Em metade das provas, as setas eram apresentadas 300ms antes da cena e na outra metade, eram apresentadas 300ms depois da cena. Ao final da exposição, era apresentado aos participantes um quadro com quatro objetos. A tarefa do participante era responder se um dos objetos estava na área apontada pela seta. Cada cena projetada possuía duas versões: uma normal e outra com partes reorganizadas. Essa segunda condição (imagem reorganizada), consistia na cena dividida em seis retângulos iguais e reorganizados formando uma nova imagem. É importante ressaltar que a orientação normal dos objetos era respeitada nessa configuração. Assim a reorganização modificava apenas a cena como um todo e não a orientação espacial dos objetos que a compunham. Os resultados sugerem que os participantes cometiam significativamente mais erros na condição de cena reorganizada em comparação ao desempenho observado nas cenas não modificadas. Esse fato foi interpretado como uma evidência da influência do esquema geral da cena na percepção dos objetos.

Para investigar a segunda questão, como acessamos rapidamente codificações de alto nível, Friedman (1979) sugere duas rotas diferentes para o reconhecimento dos objetos: a detecção de características e a análise de características. A detecção das características seria um processo eficiente, decorrente da antecipação do objeto por uma reapresentação geral da cena, nomeado esquema da cena. Nesse processo, o reconhecimento ocorreria por uma análise ampla e superficial do objeto, caracterizado pela detecção de características globais e mais salientes do estímulo. Uma ilustração de uma detecção de características seria a visão de um bloco retangular de superfície brilhante e com altura próxima a de uma pessoa em uma cena de cozinha rapidamente reconhecido como uma geladeira. Quando é encontrado algum objeto não descrito pelo esquema da cena, tem início o processo de análise de características. Essa análise seria mais lenta, detalhada e demandaria um alto engajamento da atenção. Esse processo de percepção seria próximo do descrido pela Teoria da Integração das Características (Treisman & Gelade, 1980).

Contemporaneamente ao estudo de Friedman, outro estudo seminal foi realizado por Loftus e

Mackworth (1978). Esses autores delinearam um experimento em que os participantes viam cenas com objetos representados através de linhas sem preenchimento. A informação visual era controlada de tal forma que os alvos e as cenas eram intercambiáveis. Por exemplo, um dos alvos poderia ser em um momento um polvo e no outro um trator, ambos poderiam aparecer tanto em uma cena de fazenda quanto em uma cena de fundo do mar. Assim, o alvo apresentava duas condições: ele poderia ser tanto consistente quanto inconsistente com a cena. Cada cena era apresentada por quatro segundos, sendo os movimentos oculares registrados durante a apresentação. Ao fim da apresentação, os participantes eram solicitados a realizar um teste de memória de reconhecimento. Os resultados demonstraram que o número de fixações oculares era significativamente superior sobre os objetos inconsistentes. Esses resultados também mostraram que os objetos inconsistentes tendiam a ser alvo das primeiras fixações oculares. Isso sugere que uma representação de alto nível é ativada nos estágios iniciais da visualização influenciando a seleção da informação desde os primeiros movimentos oculares.

A partir desses achados iniciais, Biederman, Mezzanotte e Rabinowitz (1982) descreveram cinco tipos de relações entre os objetos que codificam um esquema geral da cena, sendo estes: 1) Suporte: a grande maioria dos objetos repousa sobre superfícies; 2) Interposição: os objetos ocupam espaços próprios e posições definidas, não podendo ocupar concomitantemente a mesma posição; 3) Probabilidade: certos objetos são mais comumente encontrados em um ambiente do que em outro; 4) Posição: os objetos em um ambiente são alocados preferencialmente em certas posições no espaço; 5) Tamanho: os objetos apresentam tamanhos relativos constantes. Nessa concepção, algumas das relações contextuais seriam mais prontamente processadas. Inicialmente, foi suposto que o processamento das relações de Suporte e Interposição seria priorizado. Após o processamento dessas relações haveria o processamento das relações de Probabilidade e Posição. O processamento da relação de Probabilidade é prioritário em relação à Posição, pois, apesar de podermos prever, com certa acurácia, qual objeto estará contido em uma determinada cena, a sua localização na cena pode variar. Finalmente, após esta sequência de processamento, a relação de Tamanho entre os objetos seria avaliada. Essa avaliação ocorreria por último, pois não seria possível perceber alterações no tamanho de um objeto sem ter outros objetos como referência (Biederman & cols., 1982).

Para testar essas hipóteses, Biederman e cols. (1982) delinearam um experimento em que o nome de um estímulo-alvo era apresentado visualmente ao participante e permanecia disponível até que o sujeito indicasse que havia memorizado a palavra-alvo e estava pronto para iniciar a prova experimental. A prova experimental era iniciada pela apresentação de um sinal de fixação por 500ms, seguido pela apresentação de uma imagem por 150ms. Após a apresentação da imagem, uma seta era apresentada como dica espacial por 500ms. A tarefa do participante era pressionar uma determinada tecla do computador caso a indicação da seta fosse coerente com a posição do alvo, e outra tecla, caso o alvo não estivesse presente. O alvo era apresentado em duas condições: consistente ou inconsistente com a cena. Na condição de inconsistência, o alvo apresentava de uma a três violações nas relações interobjetos (Suporte, Interposição, Probabilidade, Posição e Tamanho), enquanto na condição de consistência o alvo não apresentava nenhuma violação nas relações interobjetos. A hipótese era que como as relações interobjetos teriam uma sequência específica de processamento, sendo certas relações processadas prioritariamente, a taxa de erros seria expressa em função da relação violada.

Os resultados mostraram que os sujeitos cometeram mais erros na condição de alvo inconsistente. Porém, contrariando as hipóteses iniciais, não houve diferença nas violações interobjetos. As violações também apresentaram um efeito aditivo, ou seja, três violações produziram significativamente mais erros que duas, que por sua vez tendiam a produzir mais erros que uma violação. Somado a esse fato, os autores observaram que os sujeitos respondiam mais lentamente quando havia uma violação, independentemente de qual fosse. Esses resultados sugerem que a informação semântica da cena, incluindo relações entre a cena e objetos, pode ser rapidamente acessada. Porém, não existe prioridade no processamento das formas de codificação entre as relações interobjetos e a cena, uma vez que não houve diferença significativa no efeito da resposta em função do tipo de violação na cena. Assim, um esquema global seria ativado nos momentos iniciais do processamento de cenas compostas por estímulos naturais. (Biederman & cols., 1982).

Diante desses modelos, De Graef, Christiaens e d'Ydewalle (1990) argumentam que os estímulos usados no experimento de Loftus e Mackworth (1978) eram desenhos muito simples, com poucos objetos dispersos na cena, o que poderia ter facilitado a segmentação do fundo e a seleção dos objetos semanticamente relevante. Além disso, esses autores

apontaram que a generalização de resultados obtidos em tarefas de memorização para o funcionamento geral da atenção pode ser imprecisa. Isso porque nas tarefas de memorização, os sujeitos usam ativamente a informação do contexto e buscam por objetos inesperados como estratégia para uma codificação mais eficiente das imagens. Porém, nas tarefas cotidianas, não é necessário o engajamento constante na memorização dos objetos e ambientes visualizados. Assim, De Graef e cols. (1990) propuseram que a busca visual seria um modelo mais adequado para o estudo da alocação da atenção em cenas naturais em comparação a tarefas de memorização.

Além disso, De Graef e cols. (1990) apontaram que as teorias de facilitação de percepção do objeto pela ativação do esquema não detalham o processo ou as varáveis envolvidas nesse processamento. É paradoxal como a representação de uma cena pode ser formada antes que os elementos que a constituem sejam identificados. Se a cena é caracterizada pelos elementos que a formam, como ela pode ser identificada sem que seus elementos sejam reconhecidos *a priori*? É como se fosse possível comer um bolo antes que seus ingredientes fossem misturados e cozidos. Uma questão principal dos modelos de processamento baseado no esquema, é que eles parecem presumir que os estímulos naturais apresentam uma propriedade perceptiva única, um efeito sinérgico não reproduzível com estímulos artificiais, muitas vezes definida como uma essência ou "gist" (Oliva, 2005), que facilitaria a percepção das cenas naturais.

Desta forma, De Graef e cols. (1990) propõem que o efeito de facilitação na percepção de objetos pelo contexto (Bierderman, 1972; Biederman & cols., 1982) poderia ser mais bem explicado pelo efeito de pré-ativação (*priming*). Portanto, a visualização de um determinado objeto pré-ativaria a representação de objetos semanticamente relacionados, reduzindo o limiar de informação necessário para o reconhecimento dos objetos. Assim, a influência do contexto seria um processo resultante do reconhecimento gradual dos objetos. Diferente da hipótese proposta por Friedman (1979), o processamento da informação apresentaria uma característica serial. Nesse processamento a influência da informação de alto nível seria tardia, ocorrendo somente após a integração das características e reconhecimento dos objetos.

Para testar essa hipótese, De Graef e cols. (1990) delinearam um experimento em que os participantes deveriam buscar por pseudo-objetos (figuras fechadas que não remetem a nenhum objeto existente) em cenas estruturadas com a representação de fotografias através de traços e linhas. Os movimentos oculares dos participantes foram registrados e cada cena possuía cincos versões: uma em que o objeto era consistente com a cena, e quatro em que eram inconsistentes com algum aspecto da cena (Posição, Probabilidade, Tamanho e Suporte). Os resultados sugeriram que, durante os primeiros momentos da busca, os objetos inconsistentes eram fixados na mesma frequência que os objetos consistentes, ao passo que nos momentos finais da busca, os objetos inconsistentes proporcionavam uma maior probabilidade de fixações oculares. Esses dados indicam que a informação de alto nível não exerce influência nos momentos iniciais da seleção, já que não existe diferença entre a probabilidade de fixação entre objetos consistentes e inconsistentes, o que acontece somente nos momentos tardios. Isso favorece a hipótese de facilitação mediante um processo de pré-ativação.

Algo importante a ser destacado é o fato que esses experimentos apresentaram, como ponto comum, o uso de imagens nas quais um objeto poderia ser consistente ou não com a cena. No entanto, diferiram significativamente quanto à tarefa a ser executada pelos participantes. Enquanto Loftus e Mackworth (1978) fizeram uso de uma tarefa de memória, De Graef e cols. (1990) utilizaram uma tarefa de busca visual. Portanto, uma questão importante é que os resultados observados são provenientes de paradigmas experimentais distintos, o que pode explicar a discrepância dos resultados obtidos.

Esta questão do uso de paradigmas experimentais distintos foi investigada por Henderson, Weeks e Hollingworth (1999) que delinearam dois experimentos em que foram utilizadas imagens nas quais um objeto poderia ser consistente ou inconsistente com a cena. Nesse experimento, a relação do objeto com a cena manipulada foi somente de probabilidade, por exemplo, em uma cena em que um balcão de bar era apresentado, um objeto consistente seria uma taça de cocktail sobre o balcão. Todavia, em uma cena inconsistente, a taça era substituída pela figura de um microscópio. Em um primeiro experimento, a tarefa dos participantes era memorizar a cena para um teste posterior de reconhecimento. Em um segundo experimento, os participantes eram instruídos a buscar ativamente um alvo indicado por uma palavra antes da apresentação da cena. A tarefa dos participantes era pressionar uma determinada tecla caso o alvo estivesse presente e outra caso o alvo estivesse ausente.

O resultado da tarefa de memorização, assim como o resultado obtido por Loftus e Mackworth (1978), mostrou uma frequência maior de fixação ocular sobre os objetos inconsistentes durante a apresentação. Entretanto, houve uma discrepância dos resultados em relação aos primeiros movimentos

oculares. No experimento de Loftus e Mackworth (1978), foi observada uma maior tendência em realizar os primeiros movimentos sacádicos em direção ao objeto inconsistente, o que não foi observado no experimento de Henderson e cols. (1999). Esse fato sugere que os primeiros movimentos sacádicos apresentavam a mesma probabilidade de serem direcionados tanto para um objeto consistente, quanto para um objeto inconsistente. Os resultados do procedimento de busca visual indicaram que os objetos consistentes tendem a ser fixados logo após um movimento sacádico amplo, além de serem localizados mais rapidamente. Já os objetos inconsistentes eram fixados mais tardiamente e demandavam um tempo maior para serem localizados. Esse resultado contrasta com os dados obtidos pelo experimento de busca visual de De Graef e cols. (1990), em que não existia uma maior tendência de fixação ocular sobre os objetos consistentes no início da busca.

Os resultados obtidos por Henderson e cols. (1999) contribuíram para o aprimoramento do Modelo do Mapa de Saliência (Morrison, 1984). Segundo esse modelo, um mapa de áreas-alvo potencialmente relevantes para fixações é formado em um estágio precoce do processamento da cena. Nessa situação, cada área receberia um grau de ativação compondo um mapa de ativações que direciona a atenção visual para a área do mapa com maior ativação, tornando possível a programação do movimento ocular para essa região. Depois de analisada, a ativação dessa região diminuiria significativamente e os recursos atentivos seriam direcionados para uma nova área de maior ativação. Nesse modelo, o que determinaria o grau de ativação de uma área são as informações de baixo nível, como luminância, contraste, cor, contorno, densidade e assim por diante. Somente depois das primeiras fixações, a informação de alto-nível estaria disponível para contribuir com o aumento da saliência de certas áreas semanticamente relevantes. Então, somente após um período inicial de movimentos oculares direcionados às áreas fisicamente salientes, as regiões de saliência semântica começariam a ter uma ativação maior, o que aumentaria, portanto, a probabilidade de serem fixadas. É relevante notar como o Modelo do Mapa de Saliência corrobora os modelos da busca visual que propõem um processamento tardio da informação de alto nível. Esse modelo apresenta várias características compatíveis ao modelo de pré-ativação proposto por De Graef e cols. (1990), em que a influência da informação de alto nível inicialmente não é processada, porém torna-se mais significativa com o tempo de exposição.

Mais recentemente, Gordon (2004) investigou o efeito da informação de alto nível sobre a alocação da atenção durante os primeiros instantes da visualização. Nesse estudo o autor fez uso de cenas representadas por linhas. Gordon (2004) aponta que o registro dos movimentos oculares pode ser uma medida enviesada para o estudo da atenção. Isso se deve ao fato de o movimento ocular ser uma medida indireta que quando usada para entender fenômenos que se prolongam no tempo, geralmente mostra-se confiável, porém em fenômenos que ocorrem em intervalos breves há uma dissociação entre atenção e movimento ocular. Assim, é possível que os estímulos semânticos sejam processados mais prematuramente do que demonstram os movimentos oculares. Gordon (2004) desenvolveu um estudo no qual foram apresentados, em metade das provas, objetos consistentes ao contexto da cena. No restante das provas, objetos inconsistentes foram apresentados. A tarefa do participante era identificar um estímulo-alvo ("%" ou "&") apresentado após a cena. Esses estímulos-alvo eram apresentados na localização previamente ocupada pelos objetos consistentes ou inconsistentes em relação à cena. Assim, este procedimento tinha como objetivo principal investigar a mobilização dos recursos atentivos pelos estímulos inconsistentes e consistentes. A hipótese inicial inerente a este procedimento era que os participantes apresentariam tempos de reação menores quando o estímulo alvo fosse apresentado em uma posição previamente ocupada por objetos que mobilizam recursos atentivos. O tempo de apresentação da cena foi variado e os participantes foram divididos em dois grupos. Um grupo de sujeitos realizou a tarefa com apresentações variando em 42ms, 98ms, 154ms, enquanto o segundo grupo realizou o experimento com apresentações variando em 70ms, 126ms e 182ms. Como uma segunda tarefa, no mesmo experimento e após a execução da identificação do alvo, os participantes foram solicitados a identificar o nome de um objeto presente na cena em um conjunto de alternativas apresentadas.

Como resultado, os tempos de reação para a identificação dos estímulos-alvo foram, em média, mais rápidos e precisos quando o alvo era apresentado depois de um intervalo de 40 à 70ms, em uma posição previamente ocupada por um objeto consistente. Por outro lado, depois de 150ms, a resposta dos participantes era mais rápida e precisa quando o alvo era apresentado em uma posição previamente ocupada por um objeto inconsistente. Desta forma, esses dados sugerem que a informação semântica consistente ou inconsistente influencia a alocação dos recursos atentivos em momentos distintos e precoces do processamento. O desempenho observado na segunda tarefa, na qual os participantes deveriam identificar o nome de um objeto presente na cena em um conjunto

de alternativas, evidenciou uma alta taxa de erros. Em conjunto, esse resultado pode sugerir a ocorrência de um processamento atentivo em momentos precoces que determina o contexto da cena. Esses recursos seriam alocados inicialmente em objetos consistentes, corroborando a hipótese inicial acerca do contexto da cena. Por outro lado, com o acúmulo de informação ao longo do processamento, os objetos inconsistentes podem gerar um conflito com a hipótese contextual inicial. O fato de os participantes apresentarem pouca precisão no reconhecimento dos objetos na segunda tarefa pode indicar que, apesar de haver uma identificação visual primitiva que permite a alocação de recursos atentivos, não há informação suficiente para a plena integração do objeto. Uma segunda interpretação possível é que a tarefa de identificação do alvo produz uma interferência que dificulta a retenção do objeto na memória imediata (Gordon, 2004).

Interessados na relação da informação semântica com a saliência física dos estímulos, Underwood e Foulsham (2006) propuseram um estudo composto por dois experimentos, um de memorização e outro de busca visual. Os estímulos utilizados nesse estudo foram fotografias de ambientes internos. Em cada cena foram apresentados dois objetos-alvo, um com alta saliência física e outro com baixa. Além disso, cada dupla de objetos era apresentada em duas condições. Em uma das condições, um objeto era consistente com a cena, enquanto o outro objeto não. Em outra condição a relação dos objetos com a cena se invertia. Os resultados mostraram que durante o teste de memorização, os objetos com alta saliência física tinham maior probabilidade de receber fixações iniciais. Porém, os objetos inconsistentes tinham a maior duração de fixações independente da sua saliência física. Já na tarefa de busca visual, o tempo da busca era menor na condição de objetos consistentes em relação aos objetos inconsistentes. Os objetos consistentes tinham maior probabilidade de receber fixações iniciais independentemente da saliência física.

A maior influência da saliência física na tarefa de memória, em relação à tarefa de busca visual, levou Underwood e Foulsham (2006) a concluir que aspectos distintos do processo atentivo atuam em cada tarefa. Essa diferença atentiva, segundo os pesquisadores, é gerada pelo nível de direcionamento cognitivo da tarefa. Enquanto na memorização o participante não possuía informação prévia sobre o estímulo apresentado, na tarefa de busca visual existe o conhecimento de um alvo a ser localizado. Esse conhecimento influenciaria diretamente a ativação das áreas no mapa de representações. Dessa maneira, em tarefas que exigem pouco direcionamento cognitivo, a saliência física tem mais influência sobre a atenção, enquanto em tarefas em que há mais direcionamento cognitivo ocorre uma maior influência da informação semântica.

Alguns estudos sugerem que as representações de alto nível podem ser rapidamente processadas na busca visual (Gordon, 2004; Underwood & Foulsham, 2006), porém não fornecem uma explicação precisa de como a informação de alto nível pode ser reconhecida sem que haja a integração das características físicas. Uma possível solução para esse impasse é a hipótese que a informação de baixo nível seja mediadora da informação de alto nível, ou seja, que as características físicas não precisam ser integradas para que a informação de alto nível seja processada. Certos padrões de características físicas podem ser associados a uma determinada informação de alto nível, facilitando a sua detecção (Evans & Treisman, 2005).

Essa hipótese foi investigada por Levin, Takare, Miner e Keil (2001) em situações em que as características físicas utilizadas para a discriminação entre artefatos e animais foram investigadas. Em um primeiro experimento, os participantes eram orientados a buscar por um alvo não especificado, cuja única informação disponível ao participante era a categoria a que pertencia (a imagem de um artefato ou de um animal). O alvo era apresentado entre distratores de categoria oposta (artefatos x animais). Como controle, em metade das provas, o alvo estava presente e na outra ausente. Os resultados mostraram que, para as condições de alvo presente, o custo temporal por item foi de 5ms/item na busca por artefatos entre animais e 16ms/item na busca por animais entre artefatos. Os resultados demonstram uma grande eficiência na localização do alvo em comparação às buscas realizadas por alvos sem nenhuma característica discriminadora, como estímulos artificiais em que o custo temporal de análise por item é em média de 40ms/item. Uma explicação possível para essa eficiência pode ser o fato dos artefatos possuírem traços retilíneos, enquanto as figuras de animais não. Para testar essa hipótese, os autores realizaram uma análise dos elementos retilíneos de os itens, o que permitiu o estabelecimento de um escore relacionado à quantidade de elementos retilíneos apresentados. Em seguida, tal valor foi comparado com a eficiência da localização do alvo. O resultado demonstrou que, quanto maior o escore de elementos retilíneos do artefato, mais facilmente este era localizado. Isso demonstra que a informação de baixo nível, elementos retilíneos, pode ser usada como mediadora para detecção da informação de alto nível, artefatos.

Ainda sobre essas questões, Evans e Treisman (2005) investigaram a capacidade do sistema perceptivo de detectar um conjunto de características físicas de

complexidade moderada (esboços incompletos dos objetos), usando-as para discriminar entre cenas com a presença ou ausência de um elemento-alvo, sem necessariamente integrar completamente as características físicas de um objeto. Inicialmente, Treisman e Gelade (1980) definiram o termo "característica" como um determinado valor em uma dimensão perceptual. Por exemplo, a cor vermelha seria uma característica na dimensão cor ou um triângulo seria uma característica na dimensão forma. Uma característica de complexidade moderada seria o conjunto de algumas características simples e genéricas que não constituem um objeto *per se*. Assim, características como a forma de um bico de pássaro ou suas asas, poderiam ser usadas para detectar o estímulo "ave"; da mesma maneira, um conjunto de rodas e textura metálica poderia ser utilizado na detecção do estímulo "carro", sem a necessidade de completar a integralização desses elementos na cena. Desta forma, na busca pelo alvo, o sistema visual registraria certas características pertinentes a uma categoria alvo (ex. "animais"), ativando conexões relacionadas a essa rede semântica. Tal ativação seria suficiente para gerar uma resposta de detecção de características de complexidade moderada, porém a informação não seria suficiente para a identificação completa do objeto.

Para testar essa hipótese, Evans e Treisman (2005) realizaram um conjunto de experimentos utilizando um paradigma de apresentações visuais rápidas e seriais (*rapid serial visual presentation*). Nesse procedimento, um conjunto de imagens é apresentado rapidamente (75ms por imagem) com o objetivo de evitar a integração das características, uma vez que o tempo disponível para esse processamento é bastante reduzido. Evans e Treisman (2005) solicitaram aos participantes que efetuassem uma resposta (pressionar uma determinada tecla do computador) assim que um alvo pre determinado, um animal ou um veículo, fosse detectado e outra tecla caso o alvo não fosse apresentado (provas com alvo ausente). Uma vez pressionada a tecla, a apresentação era interrompida, e então, era requisitado ao participante digitar qualquer informação sobre o alvo (por exemplo, grupos superordenados como: mamíferos, aves, répteis, anfíbios e peixes; ou alguma característica, como: cauda, bicos, asas, quatro pernas e assim por diante) e depois determinar sua localização (direito, esquerdo ou no centro da imagem). Foram utilizadas duas condições de distratores, uma com o conjunto de imagens contendo humanos e na outra contendo plantas.

Os resultados mostraram que, quando os distratores eram figuras humanas, a detecção de animais era prejudicada, porém a de veículos não. Esse resultado pode ser explicado pelo fato de animais e humanos compartilharem muitas características básicas. O mesmo não acontece com plantas, apesar das imagens frequentemente compartilharem características contextuais comuns às cenas com animais, como imagens de campos e florestas. Nesse caso, as características de plantas e animais são suficientemente distintas para uma detecção eficaz. Caso a identificação fosse completa, os participantes seriam capazes de detectar animais com a mesma eficiência observada na condição com distratores humanos e com plantas. Portanto, a interferência na detecção de animais por distratores humanos é uma evidência favorável à ideia de que as características de complexidade moderada podem mediar a discriminação de alvos. O resultado do questionamento posterior à prova revelou que os participantes eram eficazes em determinar o grupo, porém retinham poucas informações específicas sobre o alvo. Além disso, a resposta de localização não foi mais precisa do que se fosse gerada pelo acaso. O sucesso dos participantes em determinar o grupo superordenado pode advir da discriminação de características de complexidade moderada, como "bicos" e "asas" para aves, "pêlos" para mamíferos, e "pele lisa" para répteis, entre outras características. No entanto, como não há a integração da informação, não é possível para o participante realizar uma discriminação mais detalhada, ou mesmo determinar a localização do alvo. Isso, de certa forma, corrobora a hipótese inicial de que as características físicas modulam a detecção dos alvos, sem a necessidade de integração da informação (Evans & Treisman, 2005).

Outro estudo que investigou a influência da informação de baixo nível na detecção de elementos de alto nível foi realizado por Oliva e Torralba (2007). Como hipótese fundamental, os autores propuseram que a detecção de uma cena pode ser realizada de forma holística, ou seja, sem a necessidade do reconhecimento específico de objetos. Nesse modelo, o reconhecimento da cena ocorreria através da codificação de elementos da configuração espacial global, sem o processamento dos detalhes da composição. A ideia central dessa proposta é que, ao invés de ver a cena somente como um conjunto de objetos, a cena *per se*, também pode ser vista como um objeto com características físicas particulares. Assim, como animais e artefatos possuem características peculiares, cada categoria da cena possui características próprias e similares entre si. Oliva e Torralba (2007) sustentam que essas características, denominadas de "envelopes espaciais", são detectadas de maneira pré-atentiva. O envelope espacial é caracterizado pelas fronteiras e limites que constituem a imagem, como o chão, as paredes, as seções e as elevações do cenário. Por exemplo, as autoestradas geralmente são formadas

por um ambiente aberto com longas superfícies horizontais preenchidas com elementos convexos (veículos), enquanto florestas apresentam um ambiente carregado de elementos com um fundo horizontal bem estruturado (árvores) conectado por uma textura horizontal (vegetação rasteira). Assim, o envelope espacial é representado pelas relações entre linhas externas superficiais e suas propriedades internas de textura.

Para determinar quais seriam os elementos constitutivos e próprios que formam o envelope espacial de cada categoria cena, foi solicitado a 17 observadores que dividissem 81 imagens em dois grupos, depois cada grupo em dois novos grupos, até a formação de oito grupos distintos. Os participantes foram orientados a não classificar os objetos pelo seu significado semântico, mas sim com base em elementos físicos estruturais gerais. Depois de realizada a divisão, foi requisitado ao participante que relatasse qual tinha sido o critério utilizado para as divisões. A partir dessas respostas, foi gerada uma taxonomia com oito categorias, que por sua vez serviu de base para a criação de cinco propriedades categóricas inerentes ao envelope espacial, que são: Grau de Naturalidade (refere-se à quantidade de retas verticais e horizontais, sendo quanto menor a quantidade de retas, maior o grau de naturalidade); Grau de Abertura (refere-se à quantidade de elementos dispersos horizontalmente); Grau de Expansão (expansão refere-se a elementos da cena que dão ideia de perspectiva e profundidade); Grau de Aspereza (refere-se ao tamanho dos principais componentes da cena e da sua relação entre si, quanto menores e próximos, maior o grau de aspereza) e Grau de Irregularidade (refere-se a desvios do solo em relação à linha do horizonte). Em seguida, foi construído um modelo computacional capaz de identificar essas categorias. A capacidade de categorização das cenas pelo modelo foi testada comparando o resultado de suas categorizações com outras realizadas por humanos. Os resultados demonstraram um alto grau de semelhança entre as respostas, o que sugere que a discriminação de cenas pode ser realizada sem a necessidade do reconhecimento detalhado dos objetos nela contidos (Oliva & Torralba, 2007).

O rápido processamento do contexto da cena e a detecção de elementos semânticos nos momentos iniciais da visualização favorece a hipótese de uma decodificação pré-atentiva da informação. Dessa forma, as características de complexidade moderada seriam detectadas de forma paralela, assim como as características simples. Porém, no estudo de Evans e Treisman (2005), o distrator humano interfere na detecção quando o alvo é um animal. Isso indica que o processamento não ocorre em paralelo, mas sim através do engajamento da atenção. Mais recentemente, Cohen, Alvarez, e Nakayama (2011) propuseram três experimentos para analisar a relação entre o processamento atentivo e a percepção de cenas naturais. Os autores delinearam duas tarefas concorrentes que exigiam a mobilização contínua dos recursos atentivos. Em uma tarefa os participantes foram solicitados a seguir visualmente quatro entre oito pontos dinamicamente apresentados. Os oito pontos eram apresentados sobre um fundo de tela que poderia ser um fundo xadrez ou uma cena contextualizada. Em algumas provas os participantes foram solicitados a indicar quais foram os pontos rastreados; em outras, no entanto, os participantes foram questionados acerca da cena apresentada ao fundo (Experimento 1a). A segunda tarefa seguiu a mesma lógica, mas, ao invés dos pontos dinâmicos, uma sucessão rápida de letras e números era apresentada. A tarefa dos participantes era contar quantos números eram apresentados nessa sucessão de estímulos. Da mesma maneira, em algumas provas uma cena contextualizada era apresentada ao fundo e os participantes eram inquiridos sobre sua característica (Experimento 1b). A carga atentiva demandada nessas tarefas foi manipulada em outros dois experimentos (Experimentos 2 e 3). Os resultados demonstraram que a percepção da cena de fundo foi significativamente prejudicada com o aumento da demanda atentiva. Isso sugere que, apesar de bastante eficiente, o processamento de cenas naturais mobiliza recursos da atenção.

Considerações finais

Os estudos aqui sumariados são relevantes, pois expandem o estudo da atenção seletiva de forma a abarcar não somente estímulos simples, mas também cenas complexas derivadas do mundo real. Esses estudos apresentam resultados que demonstram que a informação de alto nível tem uma forte influência na alocação da atenção e seleção de estímulos. Porém, há certo desacordo acerca das etapas cognitivas envolvidos nesse processo. Vários estudos baseiam as interpretações em modelos que consideram a seleção discreta na qual a informação de alto nível é processada em momentos tardios (Duncan & Humphrey, 1989; Treisman & Gelade, 1980; Wolfe & cols., 1989). Por outro lado, outros trabalhos apresentam resultados que sugerem um modelo de processamento baseado em esquemas, que preveem uma influência da informação de alto nível em momentos precoces do processo de seleção da informação visual (Biederman, 1972; Friedman, 1979). Isso cria uma controvérsia acerca da ação da atenção em diferentes contextos que tem

suscitado um longo e prolífero debate ao longo das três últimas décadas de investigação na área. Mais recentemente, alguns trabalhos têm sugerido, de forma bastante consistente, que os recursos atentivos são importantes também na seleção e integração rápida da informação (Cohen & cols., 2011; Evans & Treisman, 2005).

Do ponto de vista prático, a compreensão desses processos é de fundamental importância para o entendimento dos comportamentos adaptativos e da seleção da informação nas mais diversas situações, desde o ato cotidiano de conduzir um veículo em segurança até a criação de sofisticados sistemas de reconhecimento de faces. Nesse sentido, vários avanços significativos foram alcançados contribuindo para a resposta de algumas questões e o surgimento de outras indagações teóricas e metodológicas sobre um aspecto fundamental da existência, a consciência humana.

Referências

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177*, 77-80.

Biederman, I., Glass, A. L., & Stacy, E. W. (1973). Searching for objects in real-world scenes. *Journal of Experimental Psychology, 97*(1), 22-27.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: detecting and juging objetcts undergoing relational violations. *Cognitive Psychology, 14*, 143-177.

Cohen, M. A., Alvarez, G. A., & Nakayama, K. (2011). Natural-scene perception requires attention. *Psychological Science, 22*(9), 1165-1172.

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effect of scene context on object identification. *Psychological Research, 52*, 317-329.

Duncan, J. & Humprey, G. W. (1989). Visual serach and stimulus similarity. *Pychological Review, 96*, 433-458.

Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: is it really attention free? *Journal of Experimental Psychology, 31*(6), 1476-1492.

Friedman, A. (1979). Framing pictures: the role of knowledge in automatized enconding and memory for gist. *Journal of Experimental Psychology: General, 108*(3), 316-355.

Gordon, D. R. (2004). Attentional allocation during the perception of scenes. *Journal of Experimental Psychology: Human Perception and Performance, 30*(4), 760-777.

Henderson, J., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*, 243-271.

Henderson, J., Weeks, P., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 210-228.

Levin, D. T., Takarae, Y., Miner, A. G., & Keil, F. (2001). Efficient visual search by category: specifying the feautures that mark the diference between artifacts and animals in preattentive vision. *Perception & Psychophysics, 63*, 676-697.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 565-572.

Morison, R. E. (1984). Manipulation of stimulus onset delay in reading: evidence for parallel programming saccades. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 667-682.

Oliva, A. (2005). Gist of the scene. Em L. Itti, G. Rees & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 251-256). San Diego, CA: Elsevier.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Science, 11*, 520-527.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Pshycology, 16*, 97-136.

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology, 59*(11), 1931-1949.

Wolfe, J. M. (2007). Guided search 4.0: current progress with a model of visual search. Em W. Gray (Ed.), *Integrated Models of Cognitive System* (pp. 99-119). Nova Iorque: Oxford.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 419-433.

Recebido em 20/10/2011
Reformulado em 09/05/2012
Aprovado em 11/06/2012

Sobre os autores:

Frederico Miranda Rodrigues Pinheiro é graduado em Psicologia pela UFU (Universidade Federal de Uberlândia), com mestrado em Psicologia pela UFU (Universidade Federal de Uberlândia). Atualmente, é psicólogo clínico da Prefeitura Municipal de Uberlândia e professor no curso de Psicologia pela UNICERP (Centro Universitário do Cerrado).

Joaquim Carlos Rossini é doutor em Psicobiologia pela Universidade de São Paulo / RP (2005) com pós-doutorado pela Concordia University, Montreal, Canadá (2010). Atualmente é professor Adjunto IV da Universidade Federal de Uberlândia atuando nos seguintes temas: processos atencionais, busca visual e psicologia cognitiva.

Contato com os autores:

Av. Pará, 1720, Bloco 2C Sala 34 – Campus Umuarama – CEP 38400-902. Uberlândia-MG.
E-mail: fredmrp@gmail.com