

Processos *bottom-up* e *top-down* no rastreamento ocular de imagens

Marcus Maia (UFRJ/CNPq)

O experimento de monitoramento ocular de imagens foi feito com 27 sujeitos expostos durante 10 segundos a versões de imagem que continham, respectivamente, elementos [-animado - humano], [+animado - humano] ou [+ animado + humano]. Cada versão podia ser precedida ou não de informação escrita presente na tela por 5 segundos, indicando o tema da imagem. Cada sujeito reportava, então, o que havia visto, registrando-se fixações e movimentos sacádicos (medidas *on-line*), bem como a reportagem final (medida *off-line*).

Este estudo procurou investigar se a existência de um título anterior à visualização da imagem é capaz de influenciar a sua exploração visual através do padrão de movimentos sacádicos e de fixações (medida *on-line*) e de reportagem final sobre o conteúdo da imagem, após a sua visualização (medida *off-line*), buscando, portanto discriminar o curso cronológico dos fatores *bottom-up* e *top-down* na percepção visual de cenas estáticas. As variáveis independentes do experimento são:

- a) Propriedades do estímulo (casal, cão, nada)
- b) Título prévio (casa, pedras, nada)

O *design* experimental cruzava os três níveis do fator (a) com os três níveis do fator (b), produzindo nove condições experimentais, cada uma das quais testada em três sujeitos, totalizando 27 sujeitos (9x3). A imagem utilizada para apresentação aos sujeitos por lapso de tempo pré-fixado em dez segundos foi o quadro “Idílio” de Tarsila do Amaral, tendo-se manipulado a imagem com o programa *Photoshop*, para gerar os três níveis da variável independente (a), ilustrados abaixo:



Figura 1: Propriedades do estímulo: cão
[+animado, -humano]

Figura 2: Propriedades do estímulo: nada [- animado, -humano]



Figura 3: Propriedades do estímulo: casal [+animado,+humano]

A variável independente (b), título prévio, consistia na apresentação por cinco segundos, anteriormente à exibição da imagem, de uma de três telas, contendo o título “Casa na colina”, “Pedras no Riacho” ou nenhum título (tela em branco).

A tarefa experimental consistia na observação da imagem seguida da redação de um parágrafo, reportando o conteúdo da imagem observada. As variáveis dependentes foram três, a saber, movimentos sacádicos iniciais, tempos totais de fixação, reportagem final. Para efeitos de análise, estabeleceram-se três regiões de interesse para as medidas *on-line*, onde foram computados os movimentos sacádicos iniciais, bem como os tempos totais de fixação, conforme ilustrado abaixo:



Regiões de interesse para as medidas *on-line*

Como se pode ver na Figura acima, as regiões de interesse eram a região onde se manipulou a variável independente (a) “propriedades do estímulo” (casal, cão, nada) e as duas regiões correspondentes a dois níveis da variável independente (b) “título prévio” (casa na colina, pedras no riacho). As duas medidas *on-line*, a saber, os índices de movimento sacádicos iniciais e os tempos totais de fixação foram computados para cada uma dessas regiões.

Utilizando-se o equipamento de **Eye Tracking**, após a calibração, o experimentador indicava que o experimento iria, de fato, começar. Após a apresentação das duas telas experimentais (5 segundos para a primeira tela e 10 segundos para a segunda tela, que continha a imagem), uma tela com a palavra FIM era apresentada, devendo, então o sujeito redigir o parágrafo de reportagem do conteúdo da imagem, conforme a instrução prévia. De modo geral, a duração total de cada sessão foi de aproximadamente 10 minutos.

Resultados e discussão

1. Figura sem elementos [+humano] ou [+animado]

Apresenta-se inicialmente o conjunto de resultados para as medidas relativas à Figura 4, que não continha nem o casal e nem o cachorro. De modo geral, esta imagem apresenta os maiores índices de interferência dos fatores *top-down*, tanto nas medidas *on-line*, quanto nas

medidas *off-line*, em relação às imagens com o casal e com o cão.

Apresenta-se inicialmente o conjunto de resultados para as medidas relativas à Figura 2, que não continha nem o casal e nem o cachorro. De modo geral, esta imagem apresenta os maiores índices de interferência dos fatores *top-down*, tanto nas medidas *on-line*, quanto nas medidas *off-line*, em relação às imagens com o casal e com o cão.

1.1. Tempos totais de fixação

Observou-se que, quando não estão presentes na imagem o casal ou o cão, os tempos totais de fixação nas regiões de interesse relativas à casa na colina e às pedras no riacho não variam significativamente entre si, na condição em que não se fornece título na primeira tela ($t=1,126$, $p=0,27$). Por outro lado, na condição em que o título é “Pedras no riacho” os tempos de fixação na região de interesse relativa às pedras é significativamente maior do que os tempos de fixação na região da casa na colina ($t=2,161$, $p=0,038$). De modo equivalente, também há diferença significativa entre os tempos totais de fixação na região das pedras e os tempos totais de fixação na região da casa na colina, na condição em que o título é “A casa na colina”. Crucialmente a diferença se dá na direção esperada, ou seja, o título determina maiores tempos de fixação na região da casa do que na região das pedras, muito embora a área física da região das pedras seja maior do que a área de interesse em que se encontra a casa ($t=4,094$, $p=0,0002$).

1.2. Percentual de movimentos sacádicos iniciais

A medida relativa aos índices de movimentos sacádicos iniciais é consistente com os achados reportados na seção acima, relativos aos tempos de fixação totais durante a visualização da imagem. Quando não estão presentes na imagem nem o casal e nem o cão, as pedras nunca são alvo das primeiras sacadas quando não se fornece título para a imagem, enquanto que a casa passa a ser objeto de 65% das sacadas iniciais. Quando o título faz alusão às pedras, os primeiros olhares são sempre direcionados à região das pedras, na imagem. De forma equivalente, quando o título se refere à casa, os movimentos sacádicos iniciais são sempre dirigidos à região da casa.

1.3. Percentual de referências na reportagem final

Medem-se, aqui, os índices percentuais de referências feitas à casa e às pedras no riacho na reportagem final (medida *off-line*), solicitada logo após a visualização da imagem sem o casal e sem o cachorro. Observe-se que, na condição sem título, a casa na colina e as pedras no riacho apresentam os mesmos percentuais de referência no parágrafo de reportagem final escrito pelos sujeitos. Por outro lado, quando o título é “Pedras no riacho”, as pedras são mencionadas em índices superiores (65%) à menção feita à casa (35%). Quando o título é “A casa na colina”, a menção à casa atinge o nível de 100% nas reportagens, enquanto as pedras ficam em 65%.

2. Figura com o elemento cão [-humano] e [+animado]

Os resultados obtidos nas três medidas, quando a figura apresentava o cão em comparação com as medidas reportadas para a condição em que a figura não exibía nem o cão e nem o humano, permitiram estabelecer a existência de efeito *top-down* resultante da ação do título prévio à visualização da imagem. Pode-se observar, agora, que a introdução do elemento cão já dificulta que este efeito se instancie de modo tão claro. De modo geral, nas três medidas, observe-se que, agora, apenas o título “Casa na colina”, relacionado a traço [+humano], apresenta efeitos significativos. O título “Pedras no riacho”, por outro lado, tem seu efeito reduzido em função da presença do cão na imagem.

2.1. Tempos totais de fixação

Os tempos totais de fixação nas três regiões de interesse não diferem significativamente entre si quando a imagem contendo o cão é exibida sem apresentação de qualquer título prévio. A comparação dois a dois revela falta de significância tanto quando se comparam os tempos totais de fixação entre o cão e as pedras

($t=0,06$, $p=0,95$), quanto entre o cão e a casa ($t=0,45$, $p=0,65$) e entre as pedras e a casa ($t=0,38$, $p=0,70$). Na condição em que o título é “Pedras no riacho”, ao contrário do que se observou na seção 1.1, em que há efeito deste título sobre os tempos de fixação na imagem que não apresenta nem o cão e nem o humano, agora já não há efeito *top-down* significativo deste título sobre os tempos de fixação nas três regiões de interesse. O cão e as pedras recebem fixações que não permitem rejeitar a hipótese nula ($t=0,79$, $p=0,43$). Igualmente não há significância na diferença de tempos totais de fixação entre o cão e a casa ($t=0,36$, $p=0,71$) ou entre as pedras e a casa ($t=0,42$, $p=0,67$). Quando o título é “Casa na colina”, também não se obtém significância na comparação entre os tempos totais de fixação recebidos pelo cão e as pedras ($t=0,12$, $p=0,89$) ou quando se compara as pedras e a casa ($t=1,58$, $p=0,12$). Significância só é obtida quando se comparam os tempos do cão com a casa, onde o título efetivamente impõe maiores tempos de fixação na região da casa ($t=2,68$, $p=0,01$).

2.2. Percentual de movimentos sacádicos iniciais

Nesta medida, pode-se observar que, de modo semelhante, na imagem em que não havia a presença do cão ou do casal, analisada em 1.2.1.2, a região das pedras no riacho continua a não atrair os primeiros movimentos sacádicos na condição em que não se apresenta título prévio. Entretanto, o mapa de elementos salientes é alterado quando o cão entra na figura, no que se refere à competição com a casa. Observe ainda nesta condição em que não se apresenta título prévio que o cão e a casa recebem os mesmos índices de sacadas iniciais, ao contrário do que foi registrado em 1.2.1.2, em que a casa na figura sem o cão ou humano recebia a maior parte das sacadas iniciais na condição sem título prévio. Quando se apresenta previamente o título “Pedras no riacho”, as pedras recebem 2/3 das sacadas iniciais e o cão fica com 1/3 desses movimentos. Entretanto, note-se que o cão impõe uma perda para as pedras, comparativamente ao registrado para a figura sem o cão ou sem o humano, analisada em 1.2.1.2, em que as pedras receberam 100% das sacadas iniciais quando o título era “Pedras no riacho”. Por outro lado, há um efeito *top-down* claro quando o título é “A casa na colina”, pois nesta condição a casa recebe 100% das sacadas iniciais a despeito da presença do cão, o que já sugere a relevância de elemento associado ao traço [+humano] no mapa de elementos salientes no *input*, pois o título “Pedras no riacho” por si só, como vimos acima, não garante a totalidade dos movimentos sacádicos iniciais.

2.3. Percentual de referências na reportagem final

Na condição sem título prévio, o cão e a casa são sempre mencionados, enquanto as pedras deixam de ser referidas em 35% das reportagens finais. Na condição em que o título é “Pedras no riacho”, as três regiões de interesse (pedras, casa e cão) recebem sempre menção em todas as reportagens. Na condição em que o título é “A casa na colina”, a casa é mencionada em 100% das reportagens, enquanto que o cão e as pedras ficam com 65% das menções, cada um.

3. Figura com elemento casal [+humano] e [+animado]

Nesta seção apresentam-se os resultados das três medidas quando a imagem exibida inclui o casal. Nessa condição fica patente o efeito do traço [+humano] no mapa de elementos salientes do input em todas as medidas. O desbalanceamento do efeito *top-down* observado na condição da imagem sem o casal e sem o cão, que já pôde ser notado na condição da imagem com o cão, é agora claramente instanciado na condição da imagem com o casal. Nessa condição, o efeito *bottom-up* do input é patente nas três medidas, com prejuízo significativo para o efeito *top-down*, ao menos nas duas medidas *online*.

3.1. Tempos totais de fixação

Na condição sem título, os tempos totais de fixação na região do casal são significativamente maiores do que os tempos observados na região das pedras ($t= 2,61$, $p=0,01$) e na região da casa ($t= 2,8$, $p=0,008$). Os tempos de fixação não diferem quando se comparam as pedras com a casa ($t=0,17$, $p=0,87$). Na condição em que o título é “Pedras no riacho”, as pedras já não recebem fixações significativamente maiores nem em relação ao casal ($t=1,44$, $p=0,16$) e nem em relação à casa ($t=1,23$,

p=0,23), diferentemente do observado na condição em que a imagem não exibia nem o casal e nem o cão, em que os tempos de fixação na região de interesse relativa às pedras foram significativamente maiores do que os tempos de fixação na região da casa na colina, por força do título “Pedras no riacho”. Agora, o efeito *top-down* do título é completamente anulado pelo efeito *bottom-up* representado pela presença do casal no mapa de elementos salientes do *input*. Também quando o título é “A casa na colina” o efeito *top-down* do título não se instancia nem em relação ao casal (t=0,75, p=0,45) e nem em relação às pedras (t=0,064, p=0,99). Note-se que o efeito *top-down* do título “Pedras no riacho” não ocorre mesmo em uma comparação com os tempos de fixação na região das pedras na condição sem título (t=2,007, p=0,054).

3.2. Percentual de movimentos sacádicos iniciais

No que se refere aos movimentos sacádicos iniciais, a presença do casal no *input* atrai sempre de forma consistente as primeiras sacadas, que nunca são dirigidas para as demais regiões de interesse, quer na condição sem título, quer nas condições em que os títulos fazem referência às pedras ou à casa. O padrão de movimento sacádico inicial em direção ao casal comprova nesta medida o predomínio do efeito *bottom-up* sobre o *topdown*.

É interessante ainda observar que a presença do casal no *input* é um atrator *bottom-up* tão influente que, mesmo em um controle realizado em que se inverteu a figura do casal para o lado direito da imagem, este elemento continuou a receber a sacada inicial no escaneamento visual da imagem.

3.3. Percentual de referências na reportagem final

Nesta condição, a referência ao casal é feita em 100% das reportagens finais, independentemente de haver título ou não, comprovando o efeito *bottom-up* já identificado nas medidas *on-line*. Registre-se, no entanto, que ao contrário das medidas *on-line*, esta medida indica efeito *top-down* do título – as pedras são mencionadas em 100% das reportagens que as mencionam como título e a casa, igualmente, é sempre mencionada quando o título a coloca em destaque.

Conclusões

Com base nos resultados obtidos, concluiu-se que:

1. A exploração visual das propriedades físicas de uma imagem (computação *bottom-up*) não é randômica. Há padrões de movimentos sacádicos e de latências de fixações relacionados a propriedades informativas ou salientes do *input*.
2. Não parece haver uma relação determinística entre a informação *top-down* e o escaneamento inicial da imagem. Os fatores *top-down* influenciam o olhar, mas podem ser sobrepujados por características intrínsecas salientes da imagem.
3. Efeitos *top-down* se instanciam mais claramente em medidas *off-line* do que em medidas *on-line*. Por outro lado, a computação *bottom-up*, capturada em medidas *on-line*, pode não se instanciar inequivocamente em medidas *off-line*.
4. O desenvolvimento de arquiteturas cognitivas que modelem a integração de algoritmos *bottom-up* e heurísticas *top-down* não pode prescindir, portanto, de pesquisa experimental sensível ao curso cronológico das operações atuantes no processo.