

# PROCESOS DE ATENCIÓN SELECTIVA PARA LA IDENTIFICACIÓN DE OBJETOS EN ESCENAS CONTEXTUALES\*

J. ARNAU; F. SALVADOR; M. PELEGRINA  
*Universidad de Barcelona*

## Resumen

La actuación de las estructuras de conocimiento en procesos de identificación de estímulos visuales ha sido operativizada frecuentemente mediante conceptos como contexto o tipicidad, los cuales pueden considerarse complementarios y producidos por la actuación de un mecanismo de tipo atencional. En el presente trabajo se intentaba comprobar si el desvío del foco de atención lograría neutralizar el efecto de tipicidad en una tarea de identificación de objetos en escenas contextuales. En los resultados obtenidos se demuestra que no se consiguió eliminar el efecto de tipicidad. Estos resultados se discuten en base al efecto producido por niveles de tipicidad de valor medio, como consecuencia de haberse anulado los valores de tipicidad extremos mediante nuestra manipulación experimental.

## Abstract

Last years many researchers have been using concepts such «contextual effect» or «typicality effect» in order to explain how people use knowledge structures for processing objects into contextual scenes. Typically effect could be considered as a focussing of attention process. In this paper we tried to prove some typicality effects related to focussing of attention process. We hypothesized that a division of the attention eliminate typicality effects. An experimental paradigm of identification objects into scenes was used. Finally, failure to accept our hypothesis was discussed into the theory of visual processing for environmental scenes.

Durante la pasada década se realizó un considerable esfuerzo investigador para explicar como los seres humanos almacenan la información sobre conocimiento del mundo en sus estructuras de memoria a largo plazo, hipotetizándose para ello diversos modelos proposicionales; por ejemplo, la teoría del esquema (Rumelhart, 1980; Rumelhart y Ortony, 1977). Algunos de estos modelos se centraron en determinar como se representa la información visual en las estructuras de conocimiento —un ejemplo de ello son los trabajos desarrollados a partir de la teoría del «frame» (Minsky, 1975)—. Posteriormente, Palmer (1977) propuso un modelo jerárquico en red para explicar como la información visual se almacena en diversos niveles de detalle. La importancia de este tipo de modelos se debe a que son «el marco en el que se insertan las nuevas teorías interpretativas de la percepción humana» (Arnaú, junio 1984, pág. 104).

Actualmente, sin embargo, una de las líneas de investigación más interesantes en percepción visual se centra en estudiar la influencia de las estructuras de conocimiento en los procesos perceptuales. La actuación de factores de tipo cognitivo en el procesamiento de los estímulos visuales ha sido señalada por numerosos autores, que hipotetizan la actuación simultánea de las características físicas del objeto y las estructuras de conocimiento del sujeto en el procesamiento de la información visual (Friedman, 1979; Prinzmetal y Millis-Wright, 1984; Rock, 1985; Virzi y Egeth, 1984). Los factores cognitivos que intervienen en los procesos perceptuales han sido operativizados mediante el uso de conceptos como «contexto» o «tipicidad» (Balota y Rayner, 1983; Flowers y Lohr, 1985; Malt y Smith, 1982; McClelland y Mozer, 1986), que pese a ser definidos de maneras muy diversas, pueden ser considerados complementarios. Así, «contexto» haría referencia a un efecto provocado por el conjunto de los elementos que forman un estímulo global y «tipicidad», por su parte, indicaría el grado de adecuación de cada elemento a la globalidad (Salvador, 1986).

\* Parte de este trabajo fue presentado como póster en la Conferencia sobre Percepción del Objeto: Estructura y Procesos, celebrada en Madrid en mayo de 1987. También queremos expresar nuestro agradecimiento a Esther Ricci y Pilar Vicente por su eficaz trabajo en el laboratorio durante el proceso de recogida de datos.

La actuación de factores de tipo cognitivo en la percepción pueden detectarse fácilmente en el procesamiento de estímulos complejos tipo escena; definiéndose ésta como un conjunto ordenado de objetos del mundo real que cumplen determinadas propiedades de carácter espacial entre ellos (Tversky y Hemenway, 1983). En diversos experimentos donde se usaron tareas de identificación de objetos en escenas se demostró que los sujetos dedican más recursos de procesamiento —tiempos de identificación superiores (Palmer, 1975) y mayor número de fijaciones oculares (Parker, 1978; Loftus y Mackworth; 1978)— a los objetos más informativos. La informatividad de un objeto hace referencia tanto a los objetos congruentes como a los objetos incongruentes con la escena, es decir, aquellos objetos que son muy típicos o muy atípicos. Por el contrario, objetos de tipicidad media poseerían un bajo componente informativo, por lo cual se les dedica un escaso número de recursos de procesamiento.

El que se efectúe el procesamiento de las escenas a partir de una rápida identificación de los objetos más informativos de la misma posibilita una gran economía de recursos cognitivos. En efecto, mediante este sistema de procesamiento identificando un escaso número de elementos la escena puede ser reconocida rápidamente y se evita una búsqueda exhaustiva de todos los objetos que la componen. La identificación de los objetos con un alto grado de informatividad puede explicarse por la actuación de unas expectativas contextuales generadas por las estructuras de conocimiento del sujeto. Dichas expectativas, que anticipan el tipo de información que será ingresada por el sujeto, provocan la actuación de unos mecanismos atencionales que dirigen el procesamiento hacia determinadas partes de la escena, aquellas donde existe una alta posibilidad de encontrar objetos informativos (Arnau y Salvador, 1986; Salvador, 1986). Así pues, es necesario un proceso de focalización atencional para llevar a cabo una correcta identificación de los elementos informativos del estímulo (Navon y Margalit, 1983; Navon y Pearl, 1985).

El presente trabajo es una ampliación de las investigaciones sobre las estrategias cognitivas que emplean los sujetos para llevar a cabo la identificación de objetos en escenas contextuales. En el experimento se efectuará una primera aproximación exploratoria al estudio de los mecanismos atencionales en la identificación de objetos dentro de escenas. Dado que, como indicábamos anteriormente, los mecanismos de focalización de la atención podrían explicarse a partir de los aspectos informativos de carácter semántico del estímulo (Butler y McKelvie, 1985; Rock y Gutman, 1981), la introducción simultánea de nuevas variables de tipo cognitivo provocarían una interferencia en la focalización atencional. En consecuencia, factores de orden semántico como la tipicidad, se verán anulados en el proceso de identificación de objetos dentro de escenas contextuales y, por ello, se

hipotetiza que factores de enmarcamiento semántico de los estímulos, tales como palabras, provocan una desviación atencional que neutraliza la actuación de las variables generadas a partir de las estructuras de conocimiento del sujeto.

## Método

### Sujetos

Participaron en el experimento 89 sujetos, estudiantes voluntarios de Psicología de la Universidad de Barcelona, de ambos sexos y con edades comprendidas entre 21 y 25 años. Todos ellos poseían visión normal o corregida mediante cristales graduados. Su participación en el experimento les permitía obtener créditos.

### Material

Los estímulos se presentaron mediante un taquistoscopio Gerbrands TA-4 de cuatro campos. Un cronómetro digital marca Letica conectado al sistema del taquistoscopio se utilizaba para medir los tiempos de reacción -TRs.

Se emplearon como estímulos un total de 90 tarjetas. De éstas, 30 tenían dibujadas escenas referidas a algunos espacios donde los sujetos realizan su actividad diaria normal —por ejemplo, «aula», «biblioteca», «comedor», «despacho», «dormitorio», etcétera—. Todas las escenas estaban formadas por un número fijo de objetos, de los cuales se dibujó su silueta sobre fondo blanco, utilizando exclusivamente líneas rectas, con trazo negro y sin sombreado.

Un segundo grupo de 30 tarjetas contenían el dibujo de un objeto situado en el centro de la misma, el cual poseía idénticas características físicas que los objetos de las escenas. Las 30 tarjetas restantes poseían una palabra escrita en el centro. Las palabras estaban también rotuladas en negro, con letra mayúscula y consistían en nombres, adverbios, adjetivos y palabras sin sentido.

### Procedimiento

La tarea de los sujetos consistió en una búsqueda de objetos dentro de las escenas, indicando si un objeto-prueba se encontraba o no en la escena presentada; según el paradigma experimental de Arnau y Salvador (1986). La secuencia de estímulos era la siguiente: primero se presentaba una tarjeta en blanco durante un período de 2.000 milisegundos; posteriormente, y durante 800 milisegundos, se presentaba un objeto dibujado y una palabra superpuestos; finalmente una tarjeta donde estaba dibujada la escena. Simultáneamente a la presentación de la tercera tarjeta se ponía en marcha el cronómetro hasta que el sujeto emitía la respuesta: «sí» cuando la escena incluía

el objeto de la segunda tarjeta y «no» en el caso contrario. Las respuestas eran verbales y un relé incorporado el taquistoscopio interrumpía el cronómetro cuando el sujeto daba la respuesta. El TR medio comprendía desde la presentación de la tercera tarjeta hasta la respuesta del sujeto. Sin embargo, si el sujeto no emitía ninguna respuesta el estímulo desaparecería del campo de visión a los 1.000 milisegundos.

Los errores fueron registrados aparte y se tuvieron en cuenta en el análisis de datos como una segunda variable dependiente. Las anticipaciones del sujeto en la respuesta —TRs menores de 400 milisegundos— o los retardos —TRs superiores a los 1.000 milisegundos— fueron contabilizados como errores. La secuencia de tarjetas se repetía 30 veces para cada sujeto, eliminándose las puntuaciones de las cinco primeras y de las cinco últimas en el análisis de datos para evitar efectos de primacía y recencia, respectivamente.

El plan experimental consistió en un diseño factorial multivariable de medidas repetidas  $2 \times 2 \times 2$ . Las variables manipuladas eran el grado de tipicidad de los objetos en cada escena, dicotomizada en dos condiciones —«tipicidad alta» y «tipicidad baja»—; la estructura física de los objetos, dicotomizada también en dos condiciones —«estructura simple» y «estructura compleja»—; y el tipo de respuesta —«sí» y «no»—. Las ocho condiciones experimentales se distribuyeron al azar para cada sujeto entre las 25 secuencias resultantes de eliminar las cinco primeras y las cinco últimas, desechándose también las secuencias restantes. Asimismo, se aleatorizaron los objetos, palabras y escenas que formaban cada condición experimental para cada uno de los sujetos, al igual que se aparejaron al azar las palabras y objetos para cada condición y sujeto.

Para obtener la complejidad física del estímulo se empleó un índice obtenido a partir del sumatorio resultante al contabilizar un punto por cada segmento del objeto dibujado, más 1/2 punto por cada ángulo no recto, más 1/4 de punto por cada discontinuidad, más 1/2 punto si el objeto se hallaba incluido dentro de otro elemento —un aparato de TV en una mesita o una carpeta encima de una mesa, por ejemplo—. Objetos con un índice menor de 15 fueron asignados a la condición «estructura simple» y objetos con un índice superior a 15 fueron asignados a la condición «estructura compleja».

El grado de tipicidad de los objetos en las escenas se midió a partir de la escala de tipicidad elaborada por Salvador y Carreras (diciembre 1984), la cual ya había sido utilizada en experimentos anteriores. Dicha escala, construida a partir de la puntuación asignada por sujetos-jueces a diversos objetos para un conjunto de escenas dadas, asocia un índice de 1 a 10 —de menor a mayor tipicidad— a los elementos de cada escena. Objetos con puntuaciones superiores a ocho fueron asignados a la condición «tipicidad alta», mientras que objetos con valores comprendidos entre cuatro y siete lo fueron a la condición «tipicidad baja».

## Resultados

Se aplicó un AMVAR —análisis multivariable de la varianza— a los datos obtenidos según el programa de ordenador P4V del paquete estadístico BMDP (Davidson y Toporek, 1983). Fue significativo el factor tipicidad con una  $F(2,87) = 21,85$  y  $p < 0,001$  en el caso multivariable;  $F(1,88) = 16,95$ ,  $p < 0,001$  para los TRs y  $F(1,88) = 18,45$ ,  $p < 0,001$  para los errores en los casos univariados. Asimismo, fueron significativas la interacción respuesta  $\times$  estructura — $F(2,87) = 4,75$  y  $p < 0,01$  para el caso multivariable, y  $F(1,88) = 9,15$  y  $p < 0,001$  para los TRs— y la interacción de los tres factores — $F(2,87) = 5,22$  y  $p < 0,001$  para el caso multivariable y  $F(1,88) = 9,06$  y  $p < 0,001$  para los TRs.

## Discusión

A partir de los resultados obtenidos rechazamos la hipótesis planteada, en el sentido de que los factores de tipicidad se verían neutralizados por un enmascarador de tipo semántico como las palabras, dado que el factor tipicidad fue el único significativo en sus efectos principales. Según se planteó en la hipótesis, la presentación simultánea de objetos y palabras tendría el efecto de dividir los recursos de procesamiento semántico, con lo cual éstos quedarían minimizados en la tarea de identificación del objeto dentro de la escena. La explicación de este efecto vendría dada por el hecho que la división del foco atencional impediría que éste dirigiera el proceso de búsqueda los objetos más informativos de la escena, por lo cual los elementos de la escena se identificarían únicamente en

Respuesta	Sí				No											
	Alta		Baja		Alta		Baja									
Tipicidad	Alta		Baja		Alta		Baja									
Estructura	Simple	Compleja	Simple	Compleja	Simple	Compleja	Simple	Compleja								
	TR	Error	TR	Error	TR	Error	TR	Error								
	759	21	711	23	698	10	714	14	719	22	760	31	708	19	718	14

Figura 1. Medias de TRs en milisegundos y porcentaje medio de errores para cada una de las condiciones experimentales.

TRs en milisegundos

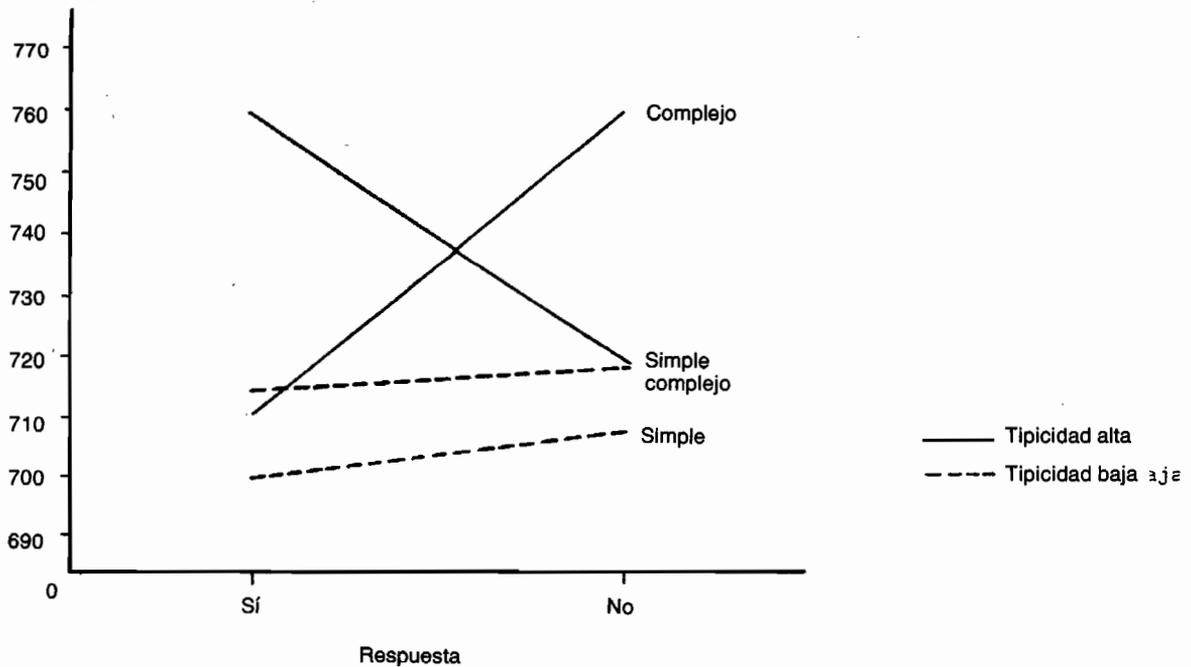


Figura 2. Gráfica de los TRs en función de la tipicidad, estructura física y respuesta de los objetos de las escenas.

base a sus características físicas. Consecuentemente, el factor de tipicidad manipulado en nuestro experimento no debería mostrarse estadísticamente significativo si queremos rechazar la hipótesis de nulidad. Sin embargo, dado que la tipicidad fue el único de los factores significativos, se invalida la hipótesis planteada y aceptamos la hipótesis nula para nuestros datos.

La condición tipicidad baja mostró TRs menores y menos errores que la condición experimental tipicidad alta. Estos resultados entran en contradicción con los obtenidos por Arnau, Carreras y Salvador (en prensa) y Salvador (1986) al presentar objetos sin máscara semántica. Aunque son necesarias nuevas evidencias empíricas, podemos hipotetizar que los resultados obtenidos se deben al efecto producido al neutralizar las tipicidades extremas. Ello favorece la identificación de los elementos con valores de tipicidad medios, los cuales se corresponden con los valores asignados a la condición tipicidad baja de nuestro experimento.

Finalmente, la interacción estructura por respuesta puede considerarse como un efecto secundario provocado por la neutralización de las variables cognitivas, que favorecen los aspectos de estructura física de los objetos. De igual forma, cabe señalar que los errores, dado el tipo de tarea requerida, se mostraron como un indicador poco sensible para medir los efectos de nuestras variables manipuladas. Estos resultados respecto los errores coinciden con los hallados en investigaciones anteriores.

Los resultados obtenidos en este trabajo deben enmarcarse en el contexto de las investigaciones sobre la actuación de factores de tipo cognitivo en el procesamiento de escenas contextuales. Con ello se pretende demostrar la actuación de las estructuras cognitivas del sujeto desde las primeras etapas del procesamiento perceptual. El presente experimento ha constituido una primera aproximación, de tipo exploratorio, al papel de los mecanismos atencionales determinados por las estructuras de conocimiento de los sujetos en la identificación de objetos en escenas.

## Referencias

- Arnau, J. (junio 1984): «Cognitivismo y percepción», *Actas de la Primera Reunión Internacional de Psicología de Tráfico y Seguridad Vial*, Valencia.
- Arnau, J., y Salvador, F. (1986): «Efectos de tipicidad en la percepción visual de escenas naturales», *Anuario de Psicología*, 35 (2), 102-114.
- Arnau, J.; Carreras, M. V., y Salvador F. (en prensa): «Bases para el desarrollo de un modelo cognitivo de la percepción visual», *Revista Latinoamericana de Psicología*.
- Balota, D. A., y Rayner, K. (1983): «Parafoveal visual information and semantic contextual constraints», *Journal of Experimental Psychology: Human Perception & Performance*, 9 (5), 726-738.
- Butler, L., y McKelvie, S. J. (1985): «Processing of form: Further evidence for the necessity of attention», *Perceptual and Motor Skills*, 61, 215-221.

- Davidson, M., y Toporek, J. (1983): «General univariate and multivariate analysis of variance and covariance, including repeated measures (URWAS)». En W. Dixon (ed.): *BMDP statistical software* (págs. 388-412), Berkeley, CA, University of California Press.
- Flowers, J. H., y Lorch, D. J. (1985): «How does familiarity affect visual research for letter strings?», *Perception & Psychophysics*, 37, 557-558.
- Friedman, A. (1979): «Framing pictures: The role of knowledge in automatized encoding and memory for gist», *Journal of Experimental Psychology: General*, 108 (3), 316-355.
- Loftus, G. R., y Mackworth, N. H. (1978): «Cognitive determinants of fixation location during picture viewing», *Journal of Experimental Psychology: Human Perception & Performance*, 4, 565-572.
- Malt, B. C., y Smith, E. E. (1982): «The role of familiarity in determining typicality», *Memory & Cognition*, 10 (1), 69-75.
- McClelland, J. L., y Mozer, M. C. (1986): «Perceptual interactions in two-word displays: Familiarity and similarity effects», *Journal of Experimental Psychology: Human Perception & Performance*, 12 (1), 18-35.
- Minsky, M. A. (1975): «A framework for representing knowledge». En P. Winston (ed.): *The psychology of computer vision* (págs. 211-277), New York, McGraw-Hill.
- Navon, D., y Margalit, B. (1983): «Allocation of attention according to informativeness in visual recognition», *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 35A (3), 497-512.
- Navon, D., y Pearl, D. (1985): «Preattentive processing or prefocal processing», *Acta Psychologica*, 60, 245-262.
- Palmer, S. E. (1975): «The effects of contextual scenes on the identification of objects», *Memory & Cognition*, 3 (5), 519-526.
- Palmer, S. E. (1977): «Hierarchical structure in perceptual representation», *Cognitive Psychology*, 9, 441-474.
- Parker, R. E. (1978): «Picture processing during recognition», *Journal of Experimental Psychology: Human Perception & Performance*, 4 (2), 284-293.
- Prinzmetal, W., y Millis-Wright, M. (1984): «Cognitive and linguistic factors affect visual feature integration», *Cognitive Psychology*, 16, 305-340.
- Rock, I. (1985): «Perception and knowledge», *Acta Psychologica*, 59, 3-22.
- Rock, I., y Gutman, D. (1981): «The effect of inattention on form perception», *Journal of Experimental Psychology: Human Perception & Performance*, 9 (2), 275-285.
- Rumelhart, D. E. (1980): «Schemata: The building blocks of cognition». En R. J. Spiro, B. C. Bruce y W. F. Brewer (eds.): *Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence and education* (págs. 33-58), Hillsdale, N.J., LEA.
- Rumelhart, D. E., y Ortony, A. (1977): «The representation of knowledge in memory». En R. C. Anderson, R. Spiro y W. E. Montague (eds.): *Schooling and the acquisition of knowledge*, Hillsdale, LEA.
- Salvador, F. (1986): *Bases per al desenvolupament d'un model d'integració sensorial i cognitiva: Efectes de tipicitat i context*, Tesis doctoral no publicada, Universidad de Barcelona.
- Salvador, F., y Carreras, M. V. (diciembre 1984): «Hacia un modelo cognitivo de la percepción visual». En M. T. Anguera (coordinador): *Procesos cognitivos: Estudios experimentales*, Simposium sobre Actividad Humana y Procesos Cognitivos, Madrid.
- Tversky, B., y Hemenway, K. (1983): «Categories of environmental scenes», *Cognitive Psychology*, 15, 121-149.
- Virzi, R. A., y Egeth, H. E. (1984): «Is meaning implicated in illusory conjunctions?», *Journal of Experimental Psychology: Human Perception & Performance*, 10 (4), 573-580.