“Seeing What You Want to See”
Top-down Effects on Low-level Vision Processes

****

University of Texas at Austin

Specific Aims

Biological vision is an amazing act, as it involves the development of a three-dimensional representation of the external world using only the firings of specific cells excited by the presence of light. From these cells, our brains are able to extract coherent patterns and organize the mass of incoming data into shapes, then objects, and from these representations develop a perception about the spatial location of the physical objects these representations depict. Considering this process, then, it can be seen that there are two main systems at work acting on vision: the external world of objects and light that present the information to us as viewers, and the internal workings of the viewers, the processing that takes place within us when we are presented with this information. Research on vision has examined each of these processes and the way that they effect the development of representations and our ability to function in the world. In particular, theories of vision can be distinguished as top-down/feedback models (e.g. Shacter & Buckner, 1998), or bottom-up/feedforward models (e.g. Brunswick & Kamiya, 1953), based on whether they argue that the brain incorporates information about the external world into its calculations, or whether it processes the information with no reference to the external world.

It has been often argued that vision is most likely a bi-directional process, with both feedback and feedforward mechanisms operating (Rao & Ballard, 1999). However, most research on feedback or top-down processing in vision has focused on high-level vision tasks (e.g. Begleiter, 1995). Low-level vision tasks such as binding line segments into lines or global contours has primarily been studied as a bottom-up process (Riesenhuber & Poggio, 2000), or as a feedback process operating at the evolutionary level rather than the individual (Geisler & Kersten, 2002). It is known that top-down processes do affect high-level vision tasks (Shacter & Buckner, 1998). This then begs the question: to what degree do top-down processes affect low-level vision? With that in mind, there are two specific aims for this proposed research.

1. To discover if the types of top-down processes based on social learning that have been discovered to affect high-level vision also affect low-level vision tasks, and if not, at which level do these processes cease to have a strong effect.
2. To discover if top-down processes which are based in explicit cognitions rather than evolved processes have any effect upon low-level vision processes.

From a research perspective, these aims are important and will advance vision research. However, they are also important from an applications perspective, as they have significant implications for the training of individuals who must make important decisions very quickly based on ambiguous visual information. The findings of this proposed research have real-world applicability.

In the remainder of this proposal, I explain the background and significance of the proposed work. I give a description of current research on both top-down and bottom-up processes in vision. I then explain the implications of the proposed research and the applications that the findings may lend themselves to. I then explain the proposed series of studies in detail, including the goals and methods of each study.

Background and Significance

Feedback models and vision

As discussed above, models of human vision initially tended to focus on one of two perspectives: top-down models that argued that the most important factor of vision was information coming in from the environment or higher-level cognitions, or bottom-up models that argued that the brain processed information coming in from the environment without receiving feedback from the world. In these bottom-up models, processing took place in the neurons, without reference to the external world. Like many fields that began with sharp binaries, a Hegelian dialectic emerged, with researchers accepting that vision is likely a bi-directional process, influenced to at least some degree by both top-down and bottom-up constraints (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; Rao & Ballard, 1999). This is reasonable, as there is research that provides support for influence from both directions.

Evidence for top-down effects in visual processing

Top-down models of visual processing argue that information about the external world affects the way that that visual representations are created. These models propose that people take into account information about the state of the external world and the relative likelihood of different representations being correct in a given context to make decisions about which possible representation is correct in ambiguous circumstances. For example, theorists arguing for top-down effects would suggest that people are more likely to see an ambiguous visual image as a monkey in the context of a jungle background than in the context of a city scene.

Theories which incorporate top-down effects are able to explain the phenomenon of priming. Priming effects occur when someone is shown a picture or given a stimulus which evokes particular concepts or ideas. If top-down effects are operating, the fact that these specific concepts have been activated should make it easier for individuals to identify related concepts in visual images. If vision is unrelated to activated cognitions and instead is a process that takes place without reference to the external world, as bottom-up theories would suggest, then there is no reason to expect priming effects.

A number of studies have in fact found priming effects in human vision (Tanaka & Sagi, 2000; Magnussen, 2000). This includes priming effects on people's capacity to recognize famous and nonfamous human faces (Begleiter, 1995), possible and impossible human actions (Verfaille & Daems, 2002), and recognition of objects (Shacter & Buckner, 1998). These findings support the idea that top-down processes affect human vision. Research by Hupe & Payne (1998) has even identified neuronal evidence that supports top-down processes in vision, as their work with monkeys has found that activation of neurons associated with higher-order cognitions affects performance on figure-ground visual tasks.

Top-down theories are characterized by an emphasis on the feedback the visual system receives from the external world. Two distinct approaches to this feedback can be identified in literature on top-down effects: theories that argue that the feedback is evolutionary, and therefore operating at the level of the species rather than the individual (e.g. Fabre-Thorpe et al 2001; Geisler, Perry, Super, & Gallogly, 2001), and theories that argue that the feedback is conceptual, and therefore determined by the individual's perspectives and existing knowledge (e.g. Shacter & Buckner, 1998). This distinction is interesting. While evolution is obviously a feedback mechanism, it operates at the level of the species rather than the individual. As a result, top-down effects caused by evolution are not amenable to change at the individual level, and in examining human vision they can only be considered something that must be similar across all individuals. They cannot be affected greatly by the idiosyncrasies of individual people.

Top-down effects which are related to existing knowledge or concepts, on the other hand, are related directly to human cognition, which is greatly affected by the social world and individual differences. Shacter & Buckner (1998) found evidence for this kind of conceptual priming, in that priming individuals with generic pictures of cars increased the speed of recognition of novel pictures of specific cars in the test that followed the prime. It can be seen that this is likely affected by the individual's familiarity and interest in cars – if someone does not have enough expertise to recognize an unusual vehicle as a car, this concept will not be activated.

This is significant, because unlike evolution, which can be assumed to be affected by the real world, human cognitions are affected by people's *perceptions* of the real world. Research in social and cognitive psychology has repeatedly found that people's perceptions of the world are heavily influenced by the social world these people exist in. For example, research on decisionmaking has shown that in making decisions, people often neglect real base rates and instead rely on heuristics based on things such as the availability of examples, or the representativeness of any given case (Tversky & Kahneman, 1974). These criteria are strongly affected by stereotypes and social learning, rather than real experience. Additionally, when presented with new information, people demonstrate bias in what they attend to and what part of the information is assimilated into their knowledge base. People are much more likely to attend to, remember, and assimilate information that is in line with existing thoughts and prejudices than contradictory information (Ross & Lepper, 1980).

If vision is affected by top-down processes related to individual cognitions and knowledge, and this knowledge is affected by prejudices and biases, then it follows that vision may be affected by these biases. People may, when presented with ambiguous visual information, see what they want to see.

Top-down effects and low-level vision

Most of the research that has been done that showed effects attributable to top-down influences has been done on very high-level visual tasks. The tasks that have been measured have mostly involved recognition of conceptual objects such as faces (Begleiter, 1995), human actions (Verfaille & Daems, 2002), and cars (Shacter & Buckner, 1998). Questions exist about the depth to which these top-down effects extend. Some researchers have even argued that research on recognition of human actions cannot be generalized to other visual tasks, as it is possible that recognition of other humans is biologically distinct from recognition of any other object (Decety & Grèzes, 1999).

Low-level vision, composed of primitive visual tasks such as the extraction of edge segments from raw visual input and the binding of these segments together, has primarily been addressed as a bottom-up process (Riesenhuber & Poggio, 2000; Brunswick & Kamiya, 1953). Some research has examined the role that top-down processes might play in low-level vision, and has found evidence to suggest that there are some feedback mechanisms that may be operating. The Bayesian model utilized and supported by Geisler and colleagues incorporates into its model an estimation of the prior probabilities of any given representation being the correct one, which can only come through feedback of some kind (Geisler et al 2001; Geisler & Kersten 2002). Research by Rao & Ballard (1999) identified neurons which seem to be optimized for the analysis of natural images, suggesting that the brain has received feedback from the natural world. There is therefore evidence that top-down effects do extend to the level of basic low-level vision tasks.

However, the types of feedback which have been examined in studies addressing top-down effects on low-level vision have all been expressly evolutionary. These theories have argued that evolution has developed in humans and other animals an enhanced capacity to process natural images, and that our visual abilities are based partially on evolved expectations about what we are seeing when we are presented with visual information. There has been no research examining the effects of expectations based on social learning or explicit information on low-level vision tasks. The proposed research addresses this gap.

<u>Implications of proposed research</u>

Many high-risk professions demand that people make significant decisions in a very short time based on visual information that may be spotty or ambiguous. Firefighters must decide if the object in the flames is a child or a blanket, policemen must decide if a suspect is holding a weapon or something harmless, and search pilots must decide if the object in the water is a man or some flotsam. The consequences of misinterpretation may be dire.

A vivid example of this type of misinterpretation is found in the case of Amadou Diallo, a Guinean immigrant who was shot to death in 1999 by four police officers who mistook the wallet he was holding for a pistol. The stimulus was ambiguous – the flat black wallet shared many characteristics with a small pistol – and the police officers mistook it for a threat. In the trial that followed, the question of racism was often raised, a charge that the officers vehemently denied. They argued that Diallo's race played no role in their actions, that they had simply mistaken what he was holding in a high-stakes environment (Duke 2000).

If it can be shown that top-down effects caused by relationships that can be learned relatively quickly do have an effect on low-level vision, rather than merely effects caused by evolutionary learning, this may show a way to rectify the positions of the officers who claim race did not play a role in their actions and those that argue it did. The officers' learned expectations may have led them to actually interpret the ambiguous stimulus as a pistol – to literally see what they expected to see. The proposed research may clarify the relationship between these expectations and socially learned stereotypes and vision, leading to implications for how to train police officers and others who must make quick decisions. Moreover, if it can be shown that these implicit stereotypes or socially learned relationships have a greater effect on vision than explicit information (as can be expected from the work of Tversky & Kahneman (1974)), then there are strong implications for training. This would suggest that it is not sufficient to give information to people relying on visual information to make quick decisions, but that it is actually necessary to train them in an environment that mirrors the real world as closely as possible. That is, it is not enough to train police officers in a situation where suspects fire upon them in 75% of the training exercises, when it is only likely to happen in 1% of real-world engagements.

<u>Proposed Experiments</u>

<u>Experiment set 1 – Low-level vision and social learning</u>

Geisler (2001) has worked extensively with low-level vision and the ability of humans to bind edge segments into contours. He has argued that this human function follows a Bayesian algorithm, which suggests that the edge segments will be bound together into a contour group when the probability, based on natural images, that they belong to the same contour is greater than the probability that they belong to separate objects. From an analysis of natural images, Geisler and colleagues identified these probabilities as a function of the edge elements distance from each other, orientation, and direction of one from the other. These three criteria provide a mathematical basis for predicting under what situations edge elements will be perceived as a continuous line.

Specifically, the Bayesian ideal observer generated by Geisler and colleagues states that the brain should interpret edge segments as belonging to the same contour C when the likelihood ratio *L,*

$$L(d,\square,\square)=\frac{p(d,\square\square|C)}{p(d,\square\square|\sim C)};$$

exceeds the criterion □,

$$\square=\frac{1-p(C)}{p(C)}$$

where □ is the ratio of the prior probability of edge segments not belonging to the same contour to their belonging. In this model, *d* represents the distance between the center of the two edge elements, □ the difference in orientation of the two edge elements, and □ the direction of one edge segment from the orientation of the second (Geisler et al 2001).

In this conception, Geisler and colleagues argue that humans have developed the calculation of □ through evolutionary feedback, making it resistant to manipulation from conceptual or cognitive effects. They have not examined this calculation, and in fact chose not to specify the value of □ in their examinations except to note that it fluctuates based on the area of analysis (Geisler et al 2001).

Experiment set 1 attempts to examine if conceptual or top-down effects can be learned which will affect performance on visual recognition tasks. If information from the environment is used in low-level vision, this will be reflected in a change in the calculations of the prior probability □, and a corresponding change in the outcome of edge-segment binding tasks. If, on the other hand, □ is set by evolutionary processes then the presence of additional information will not affect performance on these tasks. In all of the following experiments in this set and others, participants will be college students who are not colorblind.

<u>Experiment 1.1</u>

Following Geisler et al (2001), participants (n=20) will be divided into an experimental and control group, brought into the laboratory and seated at computers. A program will be used to generate line contours by linking edge segments and then manipulating the line length, degree of "jitter" of line segments, 1/*f* exponent, and the RMS amplitude of the contour. The resulting contours will be embedded in a field of similar edge segments. Geisler et al (2001) has found this to be a useful task in measuring the performance of individuals on edge segment binding, as these manipulations change *d*, □, and □ systematically.

Participants will be presented with two fields, and asked to identify which, if any, of the two fields contains a contour. Before the experimental trials, participants will be practiced in the task until they reach asymptotic performance. During the practice period, the fields that are presented will be colored either red, blue, or black. In the control condition, there will be no difference between the fields. In the experimental condition, fields shown in blue will contain a contour 75% of the time, fields shown in black will contain a contour 50% of the time, and fields shown in red will contain a contour 25% of the time.

During the experimental trials, participants will be presented with two fields and asked to identify if either of the fields contains a line, and if so which one. One field will always be black, and one in either red or blue. A contour will be present in each of the fields 50% of the time, regardless of color. If edge binding is controlled entirely by evolved constraints and is not affected by socially learned or high-level cognitions, there is no reason to expect that the additional knowledge given by the color of the field will affect processing, and the behavior of the experimental group will mirror the behavior of the control group. If, on the other hand, the presence of environmental information does affect visual processing, then it can be expected that experimental participants will make errors in favor of the blue fields and against the red.

Experiment 1.2

The second study in this series attempts to replicate experiment 1.1 with slight modifications. In this study, participants (n=20) will be divided into control and experimental groups and brought into the laboratory. As above, line contours will be generated and embedded in fields of line segments. Participants will be presented with two fields and asked to identify which of the two contains a line. During practice, the fields will be presented either in blue or red. In the control condition, there will be no difference between the blue and red fields. In the experimental condition, the red fields will always contain a horizontal contour. The blue fields will always contain a vertical line.

Once asymptotic performance has been reached, the participants will move to the experimental sessions. In these sessions, participants will be presented with two fields, both of the same color. As in the practice, fields presented in red will contain a horizontal line in one of the two fields, while those presented in blue will contain a vertical line in one of the two. If learned top-down effects do not effect low-level vision, then there should be no difference between the control and experimental groups, and both group's behavior should replicate the findings of Geisler et al (2001). If participants are utilizing the extra information given to them by the color of the line segments, on the other hand, performance should be somewhat improved over the control group.

Experiment 1.3

Study 1.3 attempts to compare the role of explicit verbal information in top-down effects on visual processing. Participants (n=20) will be brought in and divided into control experimental groups. Contours will be generated and hidden in fields of edge elements. During the practice, both the experimental and control groups will be presented with two fields, one colored blue or red and one black and while. Each color of field will contain a line segment 50% of the time in practice.

After asymptotic performance is reached, participants in the experimental condition will be presented with a sheet of paper containing information about the coming experimental trials. It will inform them (deceptively) that the color of the field will give them information about which field contains the contour. Participants will be told that when the colored field is red, it will contain a contour 75% of the time. Blue-colored fields will only have the contour 25% of the time.

Participants in both conditions will then be presented with a series of fields, presented 2 at a time. One will always be black and white, the other colored either red or blue. Regardless of color, the colored field will contain the contour 50% of the time. If explicit information about baserates does not affect vision, then the experimental group should perform no differently than the control group. If this knowledge does affect vision (or at least decisions made about ambiguous visual information), then people should favor the red fields more than the blue.

The studies presented in experiment set 1 should give us insight as to any effects that conceptual or learned top-down effects have on low-level vision processes. Contour grouping is among the most primitive visual tasks, as it is one of the most basic. Further experiments will attempt to go to a higher level.

Experiment Set 2

Experiment set 2 attempts to expand the study of contour grouping by taking it to a slightly higher level of global contours. That is, this set of experiments examines the potential effects of detection at the level of shapes, rather than just lines. Identifying these shapes requires the identification of line contours, and then the binding of these contours together. It is possible that the extra information given by people's awareness of the required relationship between line segments in specific shapes may make these tasks easier (showing top-down effects). That is, people may know that a square requires something close to a right angle, and therefore guide their attention toward these cues.

Experiment 2.1

Experiment 2.1 will replicate Geisler et al (2001) with some modifications. Participants (n=30) will be divided into two experimental and one control groups. A computer program will be used to generate lines and embed them in a field of edge segments. Participants in the control condition will be trained to identify which of two fields contain the contour, and tested on the same task.

In the experimental conditions, the program used to create line segments will be modified to generate squares, defined by having four sides of roughly equal length which meet at roughly right angles, and triangles, defined as having three sides of roughly equal length meeting at roughly equal angles. As in earlier experiments, the program will modify the size, jitter of elements, RMS amplitude of lines, and $1/f$ exponent of the line segments which make up the sides of the shape. Participants will be trained to asymptote on identification of shapes in these fields, and then tested. One group will be presented with squares, the other with triangles.

This experiment has two goals. It will provide information effects of top-down knowledge of global contours on the recognition of objects, in that the experimenters will be able to compare the performance on this task to participants' performance on line-segment recognition. However, it is acknowledged that this is confounded by the fact that squares have more components which aid in identification even at the level of edge segment binding. Geisler et al (2001) found that an angle of 90°, necessary by definition in a square, greatly increased L. The addition of triangles will allow us to compare the effects of high-level information to those caused by the shapes themselves, in that if performance is driven primarily by high-level knowledge then the performance should be similar between squares and triangles. If, on the other hand, any difference in performance between the experimental and control groups is due to the characteristics of the shapes themselves, then there should be notable differences between the square and triangle experimental groups. The secondary goal of this experiment is to provide a baseline about performance on contour recognition at the level of shapes which will then allow the experimenters to proceed with the other experiments in this set.

Experiment 2.2

Experiment 2.2 will attempt to discover if top-down processes operate at the level of global contour recognition. Participants (n=20) will be split into control and experimental groups. Shapes (squares and triangles) will be generated from edge elements and disguised in fields of edge segments as in experiment 2.1. Participants will be presented with two fields, one of which contains a shape, and asked to identify which field contains the contour.

During the training sessions, the fields that are presented will be presented in blue, or red. In the experimental condition red fields will always contain a square, and blue fields will always contain a triangle. There will be no difference in the control condition, as both colors will contain either shape.

When the participants have reached asymptotic performance, they will move onto the experimental sessions. During the experimental sessions, participants will be presented with two fields of the same color. As in the training, in the experimental condition the shapes hidden in red fields will always be squares, and the shapes in blue fields will always be triangles. There will be no difference in the control condition. If the participants are using high-level knowledge of context to guide their attention, then they should perform better in the experimental condition than in the control.

Experiment 2.3

In this study, participants (n=20) will be split into control and experimental groups and trained the same way as the control groups in experiment 2.2. They will be presented with pairs of either blue or red fields, one of which contains either a triangle or a square, and asked to identify which contained the shape. During the experimental sessions, the experimental group will be given an information sheet which informs them that in the following tests, red fields will have a square 75% of the time and a triangle 25% of the time, and blue fields will have a triangle 75% of the time and a square 25% of the time. The control condition will be given no information, and both will be tested on trials where each color has each shape 50% of the time. If the participants are using this high-level conceptual information to guide their processing, they should be more successful at identifying the squares in the red fields and the triangles in the blue fields than the control condition.

These three experiments will test to see if top-down effects exist at the somewhat higher-level task of global contour recognition, rather than just edge-segment binding. If effects are found here but not at the lower levels, they will give us information about the depth to which top-down effects resonate, and at what point the processing becomes primarily bottom-up.

Conclusion

Taken as a whole, the proposed research will examine the role of human perceptions and social knowledge in vision. This work is important, as while research on vision has accepted the likelihood of vision being a bi-directional process with both feedback and feedforward components, the research on this has focused mainly on evolved feedback mechanisms. Psychologists have demonstrated that personal beliefs have dramatic effects on a number of human processes, but vision research has been left out of this. Given the scant research on conceptual top-down effects in low-level vision, there is a starting point for this inquiry. I believe that the proposed research will address this gap in the research, and begin to examine the roles that personal beliefs and social learning play in human vision. Do people really see what they want to see? This research may answer that question.

Literature Cited

Begleiter, H., Porjesz, B., & Wang, W. (1995). Event-related brain potentials differentiate priming and recognition to familiar and unfamiliar faces. *Electroencephalography and Clinical Neurophysiology, 94*, 41-49.

Brunswick, E., & Kamiya, J. (1953). Ecological cue validity of 'proximity' and other Gestalt factors. *American Journal of Psychology, 66*, 20-32.

Decety, J., & Grèzes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Sciences*, *3,* 172–179.

Duke, L. (2000). Trial opens in N.Y. police killing. *Washington Post, Feb 3*, A15.

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience, 13*(2), 171-180.

Geisler, W., & Diehl, R. A Bayesian approach to the evolution of perceptual and cognitive systems. Manuscript under review.

Geisler, W. & Kersten, D. (2002). Illusion, perception, and Bayes. *Nature Neuroscience, 5*(6), 508-510.

Geisler, W., Perry, J., Super, B., & Gallogly, D. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research, 41*, 711-724.

Magnussen, S. (2000). Low-level memory processes in vision. *Trends in Neurosciences, 23*(6), 247-251.

Rao, R., & Ballard, D. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience 2*(1), 79-87.

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience, 3*, 1199-1204.

Ross, L, & Lepper, M. (1980). The perseverance of beliefs: Empirical and normative considerations. In R.A. Shweder & D. Fiske (Eds). *New Directions for the Methodology of Behavioral Science: Fallible Judgments in Behavioral Science Research*. San Francisco: Jossey-Bass, 17-36.

Shacter, D. & Buckner, R. (1998). Priming and the brain. *Neuron, 20*, 185-195.

Tanaka, Y., & Sagi, D. (2000). Attention and short-term memory in contrast detection. *Vision Research, 40*(9), 1089-1100.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Verfaillie, K., & Daems, A. (2002). Representing and anticipating human actions in vision. *Visual Cognition, 9*(1-2), 217-232.