

Characterising the visual system’s representations of fire

Fintan S. Nagle

October 23, 2013

1 Background

Fire is a complex and dynamic phenomenon. Characterised by rapidly shifting patterns of colour and motion, it evokes rich visual percepts as well as aesthetic responses. Over the last 1.8 million years, the evolving human visual system has been exposed to a large amount of flamelike stimuli [1] whose perception, control and mastery conferred a key evolutionary advantage[1]. We report an investigation of the visual features useful for discrimination between similar fires, the processes responsible for comparing them, and the specificity of the neural representations involved.

To engage the neural structures involved in fire perception, we used a delayed-match-to-sample task whose stimuli comprised short video clips of a hearth fire. In each trial, subjects viewed a sample clip followed by two slightly longer candidate clips, one of which was a padded version of the sample.

In such a 2-alternative forced choice task, evidence for either option is accumulated until a decision can (or must) be made[2]. We characterised the evidence present in fire clips by looking at the relationship between clip length and discrimination accuracy.

Fire clips provide different kinds of evidence: colour detail and luminance detail. Luminance-defined motion is processed better than colour-defined motion[3, 4, 5, 6], but colour-defined motion is nevertheless perceptible[7]. We measured the importance of colour-defined motion using monochromatic test stimuli.

Motion is categorised as either first- or second-order[8, 9]. First-order motion is due to displaced form (the motion in time of correlations in space), whereas second-order motion is due to dynamic contours (motion of patterns which when still are not differentiable from the background).

Attention to these patterns of motion induces activation in separate neural structures[10, 11], implying that they may be processed differently. A third type of motion percept, apparent movement, is induced by noncontinuous displacement (as in the ϕ and β phenomena)[12]. We altered the frame rates and temporal playback direction of video stimuli to examine the effects of different types of motion.

Object recognition is a key function of vision[13, 14] and is linked with outline shape[15] because outlines delineate objects’ spatial extent. To investigate whether object perception is important in fire perception, we measured subjects’ discrimination accuracy on fire clips altered to show just object edges.

The visual system is specialised for certain types of stimuli, such as faces[16, 17] or words[18]. Some stimuli provoke responses which differ in accuracy or response time[19]. We wondered whether the evolutionary benefit conferred by fire had specialised the human visual system. Expertise is commonly indicated by a performance boost when discriminating upright stimuli[?], and we found no evidence of an orientation-specific model for fire.

2 Characterising the decision process

To study the mechanisms at work in the perception of fire, we used a delayed match-to-sample 2AFC task involving short video clips of a hearth fire. In each trial, subjects viewed a sample clip followed by two slightly longer candidate clips, one of which contained the sample. The candidates were made longer due because the increased recognisability of start and end frames allowed matching to be done without attending to the rest of the sample. Subjects were asked to indicate in which candidate the sample was present, and their performance measured.

We used this matching paradigm in a series of experiments to characterise the comparison process and thus the representations involved. Matching is imperfect; we show that looking for a sample clip in longer test clips reduces performance. We then show that different types of visual information (colour, spatial form, and temporal order) are of very different importance in constructing representations of fire.

The 2AFC decision process can be seen as an accumulation of evidence supporting each choice[2]. While watching fire clips, subjects integrate evidence until a threshold is reached, allowing a decision to be made.

A percept which is present in one sample clip, but not the other, constitutes strong discriminatory evidence. If percepts from one sample are similar to those from another sample, the evidence they give is weaker and the two clips are more perceptually self-similar.

To investigate how self-similar fire clips appear, we varied the sample clip length in the matching paradigm. In Experiment 1 (see Fig. ??), sample length was varied (between blocks) between 10 and 50 frames (0.2 and 1 seconds). The ratio of sample length to candidate length was varied (within blocks) between 1.2 and 2 (candidates up to twice as long as targets).

Our prediction was that if fire clips were very differently perceived (individuated), a short sample clip would provide enough evidence to discriminate among either short or long test clips, and thus accuracy would not depend on test length. Conversely, if clips were perceived as self-similar, longer clips would induce more percepts which were erroneously matched with the sample (false evidence).

As shown in Fig. ??, it is candidate ratio (not candidate length) which has a highly significant negative effect on accuracy. It is not therefore the case that a unique, salient feature in the sample is being searched for in the test clips. There is high perceptual confusion between the sample and the false test. This means that fire clips are perceived as highly self-similar.

3 Aspects useful to fire discrimination

Whole-object representations, such as those found in human and macaque inferotemporal cortex[20, 21, 22] are built from a range of restricted visual features such as object boundaries, colour, motion, texture and contour[23, 24]. We investigated the relative importance of these ingredients in the representation and comparison of fire. The same delayed-match-to-sample paradigm was employed, with an important difference: the sample was modified to remove some of its information.

Consider showing a monochrome sample clip, with missing colour information, followed by two full-colour candidate clips. The representation set evoked by the sample is therefore lacking in colour information from the sample.

The candidate clips, on the other hand, are represented with their real colour. The process which compares candidates to samples must therefore match a detail-lacking representation to a detail-rich one. If the missing details are an important part of the representations, matching accuracy will suffer; however, the absence of unimportant details will not impact accuracy.

In Experiment 3 we removed three types of visual information from the samples: colour (by using monochrome clips), spatial arrangement (by rotating the samples through 180 degrees) and temporal arrangement (by playing the samples backwards). We measured the relative effect on accuracy of removing each detail.

Temporal inversion induced the greatest drop in accuracy (0.02, with 100% accuracy expressed as 1) followed by reversed playback (0.04). Spatial inversion produced the greatest drop (0.11, or 11%). (NB -STATISTICS WILL BE ADDED HERE)

We thus conclude that, in the set of representations accessible to the comparison process, colour is of much less importance than spatial and temporal form. The perception and recall of flame is therefore closely linked with the spatial arrangement, and the motion, of perceived image features.

4 Object-based representations

Some forms of incoming visual information is easily segmented into objects, such as faces, bodies and foodstuffs[20]. Object-centred representations allow abstraction away from retinotopic position[25] and efficient coding as object files[26, 27]. We wondered whether object representations of flames are important for fire recognition.

Objects are delimited by their edges, which describe their shape. Texture and colour information remains inside the outlined area. Recognition proceeds very well using just their edges or outlines, as shown by the easy recognition of faces from line drawings[28] and the drop in processing required to interpret line drawings[29]. We therefore hypothesised that if object recognition was important in fire discrimination, unaltered fire clips would be matched with edges-only fire clips with high accuracy. Conversely, if textural features occurring inside objects were more important, this type of cross-matching would greatly impair performance.

We repeated the altered-sample matching task, employing a Sobel edge-detection filter[30] to preserve only pixels with the highest spatial derivative.

This filter extracted the edges of flames very well, replacing their interior with a black background.

Discrimination accuracy dropped from x to x ($p=x$), showing that representations built from edges alone were matched accurately with fully-informed representations. Edge information is therefore very important for fire perception. Although we present no evidence for the existence of object files or nonretinotopic representations, the widespread ease of object recognition from outlines suggests that object files may be important for fire discrimination.

5 Motion information

Experiment 4 showed a series of clips and required a "forwards-or-backwards" judgement after each clip. Clips varied in frame rate, with interframe delays ranging from 0.02 s (50 fps, the original frame rate) to 0.2 s (5 fps). Clips were also shown rotated by 0, 90, 180 or 270 degrees (Fig. 5). Clip speed or length was not varied.

The lower we set the frame rate, the less motion information subjects could obtain from clips. Reducing frame rate degraded first- and second-order motion equally, while leaving motion percepts due to apparent motion intact.

Mean accuracy was 90% at full frame rate, but descended quickly to chance at 10 fps (interframe delay: 0.1 s). First- and second-order motion information was thus extremely important to fire perception.

Below 10 fps, any apparent motion perceived in the image is not useful for discrimination. This may be because forms in the image change shape so much over a tenth of a second that they cannot be equated, leading to no apparent motion.

Experiment (edges), by constructing edge-filtered clips, removed all second-order motion information. The small drop in performance revealed that second-order motion was not very useful for fire discrimination.

We conclude that first-order motion is of great importance in the representation and discrimination of fire.

6 Specialisation by direction tuning

Certain local neural codes are specialised: they are adapted to representing specific stimulus classes. For example, we can easily read out the presence or absence of a face by monitoring face cells in STS[31]. Some representations are tuned to stimuli oriented in a particular direction.

In face recognition, the improved recognition of upright faces over inverted faces is termed the inversion effect[32]. Although there is debate about its exact causes[?], it is generally accepted as evidence of specialisation for face processing. The representation set which informs face recognition is better tuned to encoding upright faces, and thus it cannot encode inverted faces as effectively. This impairs recognition.

The face inversion effect is present in delayed match-to-sample 2AFC tasks[33]. We checked for the presence of a similar effect in the task of fire discrimination, using two experiments.

Firstly, Experiment 3 used the same delayed-match-to-sample paradigm as Experiment 1, with one factor only: in half the trials, the sample and candidate clips were both inverted. This tested the ability of the brain's representation set to store and evaluate codings of inverted fire clips. As shown in Fig 5, there was no significant difference in recognition rates between upright and inverted clips. In fact, 50% of subjects performed slightly better with inverted clips.

Secondly, we looked at recognition accuracies across the four rotation directions used in Experiment 4: upright, inverted, and rotated 90 degrees to the left and right. As shown in Fig 5, there was once again no significant difference in accuracy across the rotation directions.

There is no large, decisive performance impairment with fire stimuli; inversion does not impair fire recognition.

This could be explained by fire representations which were direction-independent. With faces, the inversion effect seems to be due to mis-coding of configural information (faces which differ in feature configuration are recognised worse under inversion than faces which differ in feature identity[34]); this information is highly vertically asymmetric, meaning that flipped configurations are badly encoded. With flames, the lack of an inversion effect indicates that the brain is as good at representing inverted fires as upright ones; its representations are symmetric.

If the representations available to the fire comparison process were better at coding a specific direction of motion (as we could imagine, given that flaming gas usually moves upward), discrimination would suffer under rotation. However, as these encodings store enough information to allow differentiation of clips rotated by 180 degrees or by 90 degrees in either direction, we conclude that they not tuned specifically to motion in any direction.

7 Discussion

8 Conclusions

9 Materials and Methods

Stimuli used in these experiments were recorded from a hearth fire under a mixture of natural and artificial lighting. Video was captured 50 Hz using a Sony INS camera with zero CCD gain. Final cropped clips were 641 pixels high by 564 pixels wide.

Stimuli were displayed, in a darkened room, on a INS monitor with a refresh rate of 100 Hz, driven by a standard graphics card. Subjects used a chin-rest at a distance of INS and were asked not to deviate their head angle from the vertical. They were not requested to fixate.

9.1 Experiment 1

Subjects were 12 UCL undergraduates with normal (or corrected) vision.

A delayed match-to-sample paradigm was used. On each trial, subjects viewed three sequential video clips of flames separated by an ISI of INS: one sample, followed by two test clips. Test clips were equal in length and longer

than the sample. One test consisted of the sample, with short sections of video before and after it (in this clip, the sample’s start position was random). The other test was an equal-length clip from a random part of the data set which did not contain the sample.

Sample lengths consisted of 10, 25 and 50 frames (0.2, 0.5 and 1 s). For each sample length, test clip lengths were set at 1.2, 1.4, 1.6, 1.8 and 2 times sample length. Sample length was kept constant within blocks, during which test length was varied.

9.2 Experiment 2

Subjects were 12 UCL undergraduates with normal (or corrected) vision.

The same delayed match-to-sample protocol was used. Here, sample lengths were 1, 3, 6 and 12 frames (0.02, 0.06, 0.12 and 0.24 s). Test lengths were 15, 20 and 40 frames (0.3, 0.4 and 0.8 s). Test length was kept constant within blocks, during which sample length was varied.

9.3 Experiment 3

The same delayed match-to-sample protocol was used. Sample clips had a constant length of 10 frames (0.2 s) and test clips a constant length of 15 frames (0.3 s).

Test clips were displayed normally. Sample clips were altered in one of three ways. a) Colour manipulation: in hue-saturation-value (HSV) space, clips were rotated 180° around the hue axis. b) Spatial manipulation: clips were played upside down c) Temporal manipulation: clips were played backwards

Sample manipulation was varied between blocks. Each alteration was presented across 4 separate blocks, along with 4 control blocks whose samples were unmodified. The order of the set of 16 blocks was randomised.

9.4 Experiment 4

On each trial, one clip was shown. Subjects were then asked to judge whether it had been played forwards or backwards.

Clip lengths were 100 frames (2 s). Clip frame rate (not length or speed) was manipulated: interframe delays were 0.02, 0.04, 0.06, 0.08, 0.1, 0.12, 0.14, 0.16, 0.18, and 0.2 s (corresponding to frame rates of 50, 25, 16.7, 12.5, 10, 8.3, 7.14, 6.3, 5.6 and 5 Hz).

Orientation was also manipulated: clips were shown at either 0°, 90°, 180° or 270°.

Clip length was varied within blocks, whereas orientation was varied across blocks, with 2 blocks per angle. Block order was randomised. Subjects were 14 UCL undergraduates.

FIGURE 1

Figure 1: A: three progressively more detailed ways of modelling the fire comparison process. B: typical natural fire. C: typical computer-generated fire. D: one of the stimuli used in this study.

FIGURE 2

Figure 2: A: one sample clip was shown before two test clips, one of which was the sample (padded). B: For each sample length, five test lengths were used. C: classification accuracies for the 3 sample lengths. D: accuracy appears to depend on

FIGURE 3

Figure 3: On the same 2AFC task (matching a sample to one of two test clips), accuracy was measured as sample length was varied. Accuracy increased with sample length.

FIGURE 4

Figure 4: Experiment 3. A: in a delayed-match-to-sample task, sample clips were altered by colour inversion, temporal inversion, and spatial inversion. This disrupted colour information, temporal order of features and spatial layout of features. B: subjects needed to match a stored, altered percept with an incoming, unaltered clip. C: decreases in accuracy showed which types of information were most important for matching: spatial configuration, followed by temporal order.

FIGURE 5

Figure 5: Experiment 4. A: Subjects were asked to indicate whether clips were played forwards or backwards. This task requires an internal model of what real fire looks like, and the ability to compare it to incoming stimuli. B: Subjects' accuracy showed no significant effect of orientation. C: accuracy declined quickly to chance when frames were separated by 0.1 seconds or more. We conclude that subjects make use of short-range correlations between frames, and that long-range correlations are either not represented or not useful.

References

- [1] David MJS Bowman, Jennifer K Balch, Paulo Artaxo, William J Bond, Jean M Carlson, Mark A Cochrane, Carla M DAntonio, Ruth S DeFries, John C Doyle, Sandy P Harrison, et al. Fire in the earth system. *science*, 324(5926):481–484, 2009.
- [2] Rafal Bogacz, Eric Brown, Jeff Moehlis, Philip Holmes, and Jonathan D Cohen. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review*, 113(4):700, 2006.
- [3] VS Ramachandran and RL Gregory. Does colour provide an input to human motion perception? *Nature*, 1978.
- [4] Patrick Cavanagh, Christopher W Tyler, and Olga Eizner Favreau. Perceived velocity of moving chromatic gratings. *JOSA A*, 1(8):893–899, 1984.
- [5] Patrick Cabanagh and Stuart Anstis. The contribution of color to motion in normal and color-deficient observers. *Vision research*, 31(12):2109–2148, 1991.
- [6] KT Mullen and Jane C Boulton. Absence of smooth motion perception in color vision. *Vision Research*, 32(3):483–488, 1992.
- [7] Simon J Cropper and Andrew M Derington. Rapid colour-specific detection of motion in human vision. 1996.
- [8] Patrick Cavanagh and George Mather. Motion: The long and short of it. *Spatial vision*, 4(2-3):2–3, 1989.
- [9] Timothy Ledgeway and Andrew T Smith. Evidence for separate motion-detecting mechanisms for first-and second-order motion in human vision. *Vision research*, 34(20):2727–2740, 1994.
- [10] Hiroshi Ashida, Angelika Lingnau, Matthew B Wall, and Andrew T Smith. Fmri adaptation reveals separate mechanisms for first-order and second-order motion. *Journal of neurophysiology*, 97(2):1319–1325, 2007.
- [11] LUCIA M Vaina, NIKOS Makris, DAVID Kennedy, and ALAN Cowey. The selective impairment of the perception of first-order motion by unilateral cortical brain damage. *Visual Neuroscience*, 15(2):333–348, 1998.
- [12] Robert M Steinman, Zygmunt Pizlo, and Filip J Pizlo. Phi is not beta, and why wertheimers discovery launched the gestalt revolution. *Vision Research*, 40(17):2257–2264, 2000.
- [13] Denis G Pelli and Patrick Cavanagh. Object recognition: Visual crowding from a distance. *Current Biology*, 23(11):R478–R479, 2013.
- [14] James J DiCarlo, Davide Zoccolan, and Nicole C Rust. How does the brain solve visual object recognition? *Neuron*, 73(3):415–434, 2012.
- [15] William G Hayward. Effects of outline shape in object recognition. *Journal of experimental psychology: human perception and Performance*, 24(2):427, 1998.
- [16] Isabel Gauthier and Charles A Nelson. The development of face expertise. *Current opinion in neurobiology*, 11(2):219–224, 2001.
- [17] Michelle De Haan, Olivier Pascalis, and Mark H Johnson. Specialization of neural mechanisms underlying face recognition in human infants. *Journal of cognitive neuroscience*, 14(2):199–209, 2002.
- [18] Laurent Cohen and Stanislas Dehaene. Specialization within the ventral stream: the case for the visual word form area. *Neuroimage*, 22(1):466–476, 2004.
- [19] Elaine Fox, Victoria Lester, Riccardo Russo, RJ Bowles, Alessio Pichler, and Kevin Dutton. Facial expressions of emotion: Are angry faces detected more efficiently? *Cognition & Emotion*, 14(1):61–92, 2000.
- [20] Nikolaus Kriegeskorte, Marieke Mur, Douglas A Ruff, Roozbeh Kiani, Jerzy Bodurka, Hossein Esteky, Keiji Tanaka, and Peter A Bandettini. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6):1126–1141, 2008.
- [21] Chris I Baker, Marlene Behrmann, and Carl R Olson. Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature neuroscience*, 5(11):1210–1216, 2002.
- [22] David J Freedman, Maximilian Riesenhuber, Tomaso Poggio, and Earl K Miller. A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *The Journal of Neuroscience*, 23(12):5235–5246, 2003.
- [23] Kazushige Tsunoda, Yukako Yamane, Makoto Nishizaki, and Manabu Tanifuji. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature neuroscience*, 4(8):832–838, 2001.
- [24] Anitha Pasupathy and Charles E Connor. Responses to contour features in macaque area v4. *Journal of Neurophysiology*, 82(5):2490–2502, 1999.

- [25] Anitha Pasupathy and Charles E Connor. Shape representation in area v4: position-specific tuning for boundary conformation. *Journal of Neurophysiology*, 86(5):2505–2519, 2001.
- [26] Anne Treisman. The perception of features and objects. 1993.
- [27] Daniel Kahneman, Anne Treisman, and Brian J Gibbs. The reviewing of object files: Object-specific integration of information. *Cognitive psychology*, 24(2):175–219, 1992.
- [28] Vicki Bruce, Elias Hanna, Neal Dench, Pat Healey, and Mike Burton. The importance of mass in line drawings of faces. *Applied Cognitive Psychology*, 6(7):619–628, 1992.
- [29] Helmut Leder et al. Line drawings of faces reduce configural processing. *PERCEPTION-LONDON-*, 25:355–366, 1996.
- [30] Josef Kittler. On the accuracy of the sobel edge detector. *Image and Vision Computing*, 1(1):37–42, 1983.
- [31] Doris Y Tsao, Nicole Schweers, Sebastian Moeller, and Winrich A Freiwald. Patches of face-selective cortex in the macaque frontal lobe. *Nature neuroscience*, 11(8):877–879, 2008.
- [32] Tim Valentine. Upside-down faces: A review of the effect of inversion upon face recognition. *British journal of psychology*, 79(4):471–491, 1988.
- [33] Robert K Yin. Looking at upside-down faces. *Journal of experimental psychology*, 81(1):141, 1969.
- [34] Alejo Freire, Kang Lee, Lawrence A Symons, et al. The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *PERCEPTION-LONDON-*, 29(2):159–170, 2000.