

Competition and Attention in the Human Brain

Eye-tracking and single neuron recordings

in healthy controls and individuals with neurological and psychiatric disorders

Thesis by

Moran Cerf

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

May 2009

Defended May 15 2009

© 2009

Moran Cerf

All rights reserved

Author.....

Computation and Neural Systems Program

24 May 2009

Advisor.....

Christof Koch

Lois and Victor Troendle Professor of Cognitive and Behavioral Biology

Advisory Committee

.....

Ralph Adolphs

Bren Professor of Psychology and Neuroscience and Professor of Biology

.....

Shinsuke Shimojo

Gordon and Betty Moore Professor of Computation and Neural Systems and Electrical
Engineering

.....

Doris Tsao

.....

Itzhak Fried

Director of the Adult Epilepsy Surgery Program, UCLA

Dedicated to my friend and mentor

Yoram Kaniuk

who taught me that film, fiction, poetry, or a Ph.D. thesis

are all the same

“You sit, and you write!”

And to my children

Acknowledgment

All I did throughout the time here is channeling other people's thoughts, ideas, inspiration, and criticism into an organized work. I chose the font for this thesis by myself. Without these people — none of this would have ever happened.

Collaborators

First and foremost I want to thank the patients who participated in my studies. Not only did they help me leave a small dent on the advance of human knowledge, but they forever left a remarkable memory with me. I remember each and every one of them, and the talks we had. I remember their preferred food, their preferred movies or TV shows, their football team preferences, and the stories they were willing to share. I took them with me to whichever conference I went to, and they are still the grounding essence of my work. My will to help them, improve their lives, or enhance their understanding of conditions they suffer from or situations in life they don't understand that seem trivial to us are the driving force behind my work. If anything I did is worth something — it is thanks to their dedication to my success.

I wish to thank Laurent Itti, Rob Peters, and Daniel Cleary for collaborating with me and tolerating me through my time with them. A special thanks is due to Wolfgang Einhäuser and Jonathan Harel who were more than just collaborators to me, but shaped forever the ways by which I do science. Wolfgang's dedication and persistence, accuracy and thrill for research are remarkable and I can only hope to be as good a scientist as he. Jonathan's diversity and ability to fear not any new idea I threw, as well as remarkable speed and creativity, inspired me to do

well. Our times together, working and discussing my projects, as well as his fierce criticism of some terrible ideas I had, which I strongly argued for, helped me stay focused and reach the goal line. I am grateful for him for that.

Nikhil Thiruvengadam, Michael MacKay, Paxon Frady, Sami Zerrade and Alex Huth were allegedly my students. I was supposed to mentor them and guide them in the ways of science. This could not be a bigger illusion. These four remarkable students are far better than me in most things related to science, and I was lucky enough to be a bit more experienced, otherwise I would be embarrassed to even sign my name on any project I collaborated with these kids on. They are the brains behind most of what I did.

Maria Moon, an incredible graphic artist, collaborated with me throughout the years and assisted me in finding new methods to present my data in ways that will make it clearer and more engaging to the reader. Her footprint is on many of the following figures.

Brad Duchaine was kind enough to help us recruit the Prosopagnosia subjects and put his trust in our research without ever meeting me in person.

Irene Wainwright and Brooke Salaz have provided valuable and prompt help throughout these years in several of the fundamental details that keep things going at UCLA.

Among my collaborators at UCLA, Matias Ison, Sasha Kraskov, Rodrigo Quian-Quira, Hagar Gelbard-Sagiv, and Roy Mukamel were an invaluable asset to my work. Sasha and Rodrigo dedicated time and efforts to helping me shape my work and make it into a clear paradigm worth of the patients' time.

Anna Postolova, Neelroop Parikshak, and Vanessa Isiaka made the time at UCLA pleasant and were as efficient and involved in a research as one can be. They helped me get through all the hurdles faced by a visitor in a foreign environment like the medical center in Westwood and made my work at UCLA fruitful.

My collaboration with Ralph Adolphs' lab had exposed me to some of his wonderful lab members. Whether it's the engaging conversations with Lynn Paul, who graciously let me get a glimpse into the world of psychiatry by letting me attend her interviews with patients, which were among times the most exciting times I had at Caltech, or through the countless hours I spent talking to Lynn about disorders and social affect on patients which are still the key ingredient in my scientific interests. Catherine Holcomb and Jessica Levine were able to make a much-disorganized being like myself into a very decent host for the autism and AgCC subjects. Dan Kennedy shared his endless knowledge on research with autism population with me and helped shape my final results.

I could not have had bigger shoes to get into than those of Gabriel Kreiman. Choosing to follow up on a project that he started as a graduate student created high expectations that I had absolutely no chance of meeting. Luckily, I had some help that Gabriel didn't have when he wrote his thesis – I had the support of Gabriel Kreiman. His thesis inspired me throughout my years of research and served as a basis for much of this dissertation. I referred to Gabriel at times when the hardship of research took its toll on me. Gabriel always had a wise and witty way to place me back on my feet. Besides Gabriel I received much support and was engaged in long discussions with Ueli Rutishauser to whom I owe great thanks.

My peers in the Klab throughout the years had a major impact on my sanity and interest in science. Our Friday morning lab meetings were the place I got the best and the most honest comments about my work that made my output improve. I enjoyed my hours there immensely and knew that my “second home” is a place of pure genius. Special thanks to Julien Dubois, Milica Milosavljevic, and Heather Hein who facilitated my work in ways that most graduate students only dream of – as a checkpoint between me and any means of bureaucracy that she took on herself to not burden me with. Teresita Legaspi of the registrar office and the entire administration of Caltech were amazingly efficient and useful throughout my studies. Having spent 4 years in the U.S. in the course of my work, and encountering numerous service providers that were supposedly doing their best to assist but in fact were repeatedly found to be incompetent I was in awe again and again when I witnessed the efficiency and the kindness of everyone I had to work with at Caltech. I wish other organizations would be able to learn from Caltech’s administration.

Einstein had Michaela Beso. Simon had Garfunkel. Bush had Dick Cheney. Bonnie had Clyde. Laurel had Hardy. And I had Florian. No counterpart had ever contributed so much to my research, put me in my place when needed, or looked at my work in the most critical way. Florian encouraged me when things did not seem to work, and made any effort to stop me when I lost my humility. Florian is the best collaborator I could ever wish for. It took us little time to figure each other quirks and methods of work and within days we were a solid team with a clear and strong agenda. If my future careers promise half the team the two of us made – I am bound to succeed. I am grateful to Florian for being by me as a scientist, a collaborator and a friend.

My thesis committee, composed of Ralph Adolphs, Shim Shimojo, and Doris Tsao (in addition to Itzhak Fried and Christof Koch) provided invaluable feedback and made my journey exciting as well as increasingly fruitful. Ralph Adolphs, besides being an avid member of my committee, also spent much of his time and energy in discussions with me that go beyond science. Some of the conversations we had, on deserted islands miles from Caltech shaped my perception of science more than he can imagine and helped me make up my mind about my deciding how to pursue my career beyond this point.

Very few people are honored to work and interact with people they admire. People they had heard of before and asked themselves what it would be like to interact with them daily. I had the fortune to partake in this experience while working closely with Doctor Itzhak Fried. A fellow Israeli whose interests go way beyond science. Book recommendations, and analogies from the worlds of classical music to marathon running were a commonality in my interactions with Itzhak. The joy of speaking with a person whose eyes lit when he sees the results of a study you conducted a couple of days before is endless. Itzhak was an inspiring part of my thesis and I am grateful for his belief in me. I hope I returned the investment.

Money, sadly, is still a required ingredient for a thorough scientific research. Throughout my years at Caltech I was funded by numerous agencies, grants, and supporters. Some of which I wish I had not received any funding from, like the DARPA agency, and their underwriting of the Lockheed Martin corporation, and some which were purely interested in promoting knowledge and advancing science. These are the National Geospatial-Intelligence Agency

(NGA), the National Science Foundation, the Mathers Foundation, and the National Institute of Mental Health.

And finally, Caltech. I have spent time at various institutions in my life. Structured places with code and guidelines. None of these even remotely compares to the California Institute of Technology. This gem in the outskirts of Los Angeles is the ultimate place for knowledge to be transferred. I have solely good memories of this place that I had never imagined existed. A place of pure honesty, critical thinking and noble men and women who care to make the world a better place. Every morning I set foot near the Beckman Institute's "gene pool" I was filled with pride and humility to be part of a unique place that set the standards for all the others.

Friends

My friends Ady Mor, Asya and Danny Rolls, Daniel Mendelman, Danny Braunstein, Ronnen Sigal, Eran Socher, Etan Ilfeld, Galit Levin, Guy Bar'ely, Ilan Danoch, Inbal Amirav, Judy Weinberger, Jonathan Sternberg, Keren Kohen, Lior Chefetz, Yoni Peltzman, Ori Heffetz, Ronit and Israel Shapira, Sharon Shochat, Roy Rozental, Ilana Neshet, Yonatan Even, Yaeli Moratt, Yair Bartal, and Yaniv Konchitchki all supported me during my work.

Guy Hoffman, Guy Ross, and Dory Lobel, in turn, each served as my closest friends throughout my years in the U.S. and shared countless hours listening to my talking, complaining, or bragging about my work. Eventually, they probably are the ones who know most about the hurdles of my research – which is an aspect of the research somewhat gone when the shiny thesis is all done....

The Rubinim – Gil Appel, Tomer Avni, Zorik Tredler, and Yonatan Rubin made sense and reminded me who I was when things were hard. Rubin was the major driving force in making me pursue neuroscience in the first place, when his stories about the program in the Hebrew University and my friendship with him inspired me to apply for a Ph.D. program. In the first year of studies and qualifying exams I relied solely on his help to get through the hardship of the first steps in neuroscience. He has been a great friend and the silent sixth member of my thesis committee.

Friends help you move, but great friends help you move bodies. My childhood trio Ben Artzi, Or Heller, and Yadin Kaplansky were the ones I looked to whenever I did something good, hoping to hear them suggest that they might be proud – or that they might care. Yadin and Or came to visit me in Los Angeles when my study was starting to shape and take form and they are the main models in my dataset of nearly a 1000 images. Each of my subjects and every user of my faces database is very familiar with their looks. I promised them then that every paper of mine will include a picture of them, as an example of my dataset images, and – although they probably could not care less – I lived up to my promise. Yadin is leading a spiritual life, believing that after years of holistic medical studies he will be able to cure cancer by interacting with people's past incarnation. He could not hate my consciousness research more. The idea of someone believing that there is neither a soul, nor a god, for that matter, repulses him. I always referred to him when I needed perspective about my work. Ben is a musician and an artist. He never was good in math or in any aspect of schooling for that matter. Having to explain to him repeatedly what I did in very simple words was a crucial step in my writing papers throughout the years. If Ben got it – I found a way to explain it to the layman. Or is a cynical person. Working as a journalist in one of Israel's leading news channels makes him feel

that science is usually a waste of time. If I could convince him that something of what I do is meaningful – I was proud. After knowing me since I was 5 years old, and staying my friend for all these years, upon my meeting Or a year towards my end of a Ph.D. in Neuroscience, Or introduced me to a person we both met as “My friend who studies Physics”. Those are my friends. Always keeping me in perspective.

Galit, who used to be my girlfriend prior to my leaving for Caltech was one of the 3 driving forces for my coming here. When she said, when we were still a couple, that she always wanted to go live a year or so abroad I started applying. When she dumped me months later, I wanted to be as far away as possible from her, so I would learn to cope with the despair of not being with the girl I loved so much. California seemed far enough. Two years later she joined me in the U.S. No words can describe the relationship I have with Galit. Something between an older sister and a closer-than-family friend. Although my study with patient 394 was one of the most remarkable ones, and where I first was able to launch my real-time spike detection system that most of my work was later based on, I will never forgive myself for choosing to stay in the summer of 2006 to test with this patient instead of standing up to my promise to fly to Israel and attend her wedding. Galit gave me stability in life, the security one needs to take a leap of faith and go do something he never thought he can, like study neuroscience and pursuing a career in science. When I read Einstein’s recent biography by Walter Issacson he mention that Einstein promised to give half of the prize money from winning the Nobel prize – if he ever did – to his first wife (who allegedly was not only a driving force in his work, but had a share in his marvelous ideas for which she never was credited), Mileva Malic. Although the terms that led to this decision were almost contrary to the ones which led me, I think that if my research ever yields anything fruitful – Galit should have it all.

In loco parentis

A few people assumed the position of parents in the lack of immediate relatives in Los Angeles, giving me the feeling of home away from home. Rachel Stull shared her vineyard in Santa Ynez with me, where some of the greatest ideas were conceived. Rachel Vert and Chaim Meital acted as my de-facto parents on a day-to-day basis and my gratitude is forever granted to them. Alon and Elina Pnini put me in my place more than once when I needed focus, and were among the family away from home, too. Paula Thompson and Michael Laird literally opened their home and treated me like a son during the course of my writing this work.

People

A few people had an influence on my thesis, direct or indirect, that I should acknowledge separately. The first is my high-school math and physics teacher: Yoav Yaron. Yoav intrigued us in Physics classes with stories and knowledge way beyond the ordinary curriculum. He inspired us to feel that science is by far the most exciting and intriguing thing one can do with one's life, and – though it took me a couple of years to apply this – I still hold him responsible for me choosing this path. Yoav also inspired us to think that teaching is one of the most exciting and important things one can do with his life. Thanks to Yoav I polished for years my skills as a teacher and a performer and feel comfortable standing in front of large audiences and talking about research. I hope to be as good a teacher as he was.

Yoav alone is not solely responsible for my love for audience and my ability to happily not fear talking about research in front of them. My years in the school of arts surely had their impact on my love for public speaking, and probably indirectly the choice of this topic for my research

– attention. Any kid who was raised to play music, dance ballet, sing, and act in front of large audiences wants attention, and, given the option, might want to even research it.

Josh Shachar, who saw beyond the immediate problems of hiring a graduate student whose mind is mostly in his research to work on real-life problems involving brain cancer, reminded me that the reasons I went to science in the first place were noble and had to do with helping people. During the research you get immersed in the questions you care about so much that you forget what it's all for. Josh had me work on a patent for a device that will ease the life of patients with brain cancer. The hours I spent in his company and the usage of my research methods and knowledge for real-life immediate problems reminded me that behind all of these questions and many papers and words describing the brain, lie people with actual need for that knowledge. I would also forever be indebted to Josh and the entire team of Pharmacokinetics for reminding me how boring, meaningless, stupid and unlawful life could be in the industry and helping me shape my decision to choose the path of a scholar.

Prof. Daniel Kahneman gave a talk in Tel-Aviv University in the summer of 2002, soon after he received the Nobel Prize for Economics. A short email correspondence with him soon after, his belief in a young scientist whom he had never met in person at the time, and his recommendation to only look at Caltech as an option for my endeavors was a very influential step in making me leap towards this research. One of Kahneman's first comments was: If you ever want to research these questions you ask me about, you had better start from the bottom by thoroughly researching something else that will give you a good background. I found his short comment of high importance and started enhancing my knowledge. This pushed me in

the right direction in science. Instead of starting from the karate fight — I decided it might be smart to first paint a fence.

Few bosses in your life stay with you long after the company is no longer part of your routine — as Amichai Shulman and Shlomo Kramer. Amichai was my manager at iMPERVA — a security company for whom I worked prior to leaving for Caltech. Shlomo was the CEO of iMPERVA. While Shlomo did not see eye-to-eye with me with regards to the importance of my research, and did not support my leaving a successful company where I got to break into banks and institutes on a daily basis, many of my methods of work and my management skills that were used when I had my own students to mentor or large projects to handle were an exact replica of the skillset and attitude I acquired from watching a world-leading manager do his job. Amichai is a relaxed and accurate manager. He calculates his steps profoundly and is the rare combination of a scientist and a manager in an industrial company. His years in academia, combined with his years in the high-tech industry, make a unique manager that I kept turning to for advice throughout my years at Caltech. Amichai and Shlomo taught me that the best way to manage — yourself or others — is through understanding of the character and the person behind the employee. Everyone has his own methods of work — and a good manager finds a way to bring the best from a given set of skills that are already there. I apply their ways of working with people daily.

As much as the support of grants was needed for my research to happen, I also needed money to live my life, as a Ph.D. is useless if not combined with inspiring ideas that can be generated only outside of the box. The American Film Institute provided me the extra “out of the box”

thinking about science, as well as much needed extra cash to support my “inspiring” times at Caltech. Working, throughout my entire time at Caltech, with screen writers and directors from AFI – writing scripts about science, scripts that portrayed science in an accurate way, having to explain research methods and ideas to people who never set foot in a lab, and having to write a clear and exciting story from my life as a scientist numerous times made my research better. I am grateful to the American Film Institute, and in particular to Christ Schwartz, Joe Petricca, and Joan Horvath who gave me this opportunity. The directors and writers Beau Tipton (“Teen Sniper”), Benjamin Epps (“Stereolife”), Dylan Tuccillo (“Hyperesthesia”), Gregory Shull (“Alpha Man”), Marc Abraham (“Flash of Genius”)¹, Jonathan Green (“Sleep”), and Karine Nissim (“In another life”) were all inspiring to me throughout my years. I would also like to extend my gratitude to the Sloan foundation who facilitated this collaboration through their funding.

Los Angeles is filled with unique architecture. Most world-renowned architects reside in Los Angeles, have their headquarters here, or at least have a significant monumental work set here. And as architects famously say: it’s all about location, location, location. A thesis doesn’t just happen in someone’s mind. It needs a venue to be conceived. My venues are clearly set in every line and sentence put in this work. I spent the majority of my time writing and thinking in one of three places: The Standard hotel on Sunset Boulevard, the Chateau Marmont on Sunset, and on the premises of Rodney Manor in Los Feliz. Each of these places had a remarkable effect on my mood and thus on my content. The gorgeous girls asking questions

¹ <http://blog.afi.com/afifest/index.php/2008/11/08/we-told-you-so-scientific-disasters-in-film-as-entertainment-or-cautionary-tale/>

while a random guy sat and wrote science papers where others hit on them cleared my mind next to the Standard pool. The famous actors who occasionally wandered into the Marmont while I was writing about showing pictures of them to my patient (and, specifically Heather Graham, who showed extreme interest in my research – which she was somewhat a subject of), and the lifestyle of Los Angeles made me and my research better.

Two people served as my mentors and advisors in my early stages of research – at the Tel-Aviv University and the Hebrew University. Those are Prof. (Major General) Itzhak Ben-Israel and Dr. Avshalom Elitzur. Each in his own way showed me the beauty of science, the beauty of research, and convinced me that answering the bigger questions of the world is something that even simple people can do. While in my mind science can only be done by people of Newton or Descartes' caliber – which I am nowhere near – Avshalom and Itzik showed me that if you tackle the questions in simple small steps and let curiosity rather than fear guide you, you might end up answering the questions you care about... without even noticing you did. And that is the beauty of science.

Three decisions I had made in my life without knowing exactly why proved to be most valuable at times during my research. The first, the choice to take classes at the Art Center College of Design in Pasadena. On top of the load of work a Caltech graduate student has, I chose to take art classes every term alongside my studies. While at first it was hard to convince Christof why learning Maya or working with InDesign is beneficial for my work, in the end I think it turned out to be one of the smartest investments I made in my academic career. The methods I learned there proved to be useful when I needed to design a computer game for my patients, create a complex figure for my work, or edit movies that would end up being a tool to

demonstrate my work to the audience. I feel I could not have made a smarter choice. My second choice was to live far from Caltech, in the exciting Los Feliz. On top of it proving to make my life much more interesting, which I feel is extremely beneficial for a scientist, and filling my days with friends and ideas coming “out of the box”, it enabled me to spend at least 40 minutes a day (20 minutes in each direction) just driving from/to Caltech and thinking. Those 40 minutes a day before and after my time in the lab proved to be the most important time for ideas to sink and work to be structured. I, since then, have recommended to every starting graduate student who asked me to do the same. Third is a choice I had to fight my parents for hours for when I was 7 years old. At the time, for my second A+ score in second grade I got from my parents this box that no other kid had heard of at the time – a computer. A PC XT with 512kb memory and no bootable operating system or hard drive was my best friend at times. My parents tried to limit my play time to a few hours a week. But I snuck minutes and seconds whenever I could. Playing the immortal “King’s Quest” computer game. This game by Sierra Online, the first company I promised myself I would work for if I ever became a grown-up, asked the player to type commands to his character. I remember myself sitting for hours with an English-Hebrew dictionary learning words like “Open Door”, or “Bow King” (the King’s Quest series wasn’t big on prepositions...). Despite my countless fights about this with my parents who felt their kid would do better job in school if he pretended to read books rather than play computer games, these hours in front of the computer were the most fruitful tutor for my English. By age 10 I could read and write English and by age 16 I could read Shakespeare. I promise myself to let my kids play as much as they need if it will end up enabling them to write a six-hundred page thesis easily without feeling worried of using a language that is not their mother tongue.

One more person who had an important role in my deciding to leave a stable life in Israel for a very unclear and unstable future is the renowned scientist Richard Feynman. His book “Surely you’re joking” to be exact. This book I read cover-to-cover on a Saturday evening, sitting on a bench on Rothschild avenue in Tel-Aviv. His inspiring stories and tales about life as a scientist and a person in Los Angeles made me realize that there are some cool scientists, and that I could actually find an exciting life moving to L.A., which eventually was the push I needed.

If thou of fortune be bereft

And in thy store there be but left

Two loaves – sell one, and with the dole

Buy hyacinths to feed thy soul.”

Six people were my hyacinths throughout the 4 years at Caltech. Yarin, my brother. Tal. Michel and Sara, my parents. And Kelsey and Ginger.

When I left Israel for the U.S. my mother cried heavily at the airport. My father briefly told me a story of his leaving his homeland for Israel to join the army. As he left the relationship between him and his dad deteriorated as they did not get to talk daily. “Even if you have nothing to say, just speaking briefly for a few minutes a day makes a world of a difference.” We had agreed, my parents and I, to talk daily on the phone. For 4 years, rain or shine, every morning or evening started or ended with a brief update between me and my parents. The daily update kept me going in hard times and was the main reason and encouragement to do what I do. Whenever someone asked me why I was doing this Ph.D. I said: “My mother needs a plaque for her restroom”. And being a Jewish mother, I am not sure how far I am from the

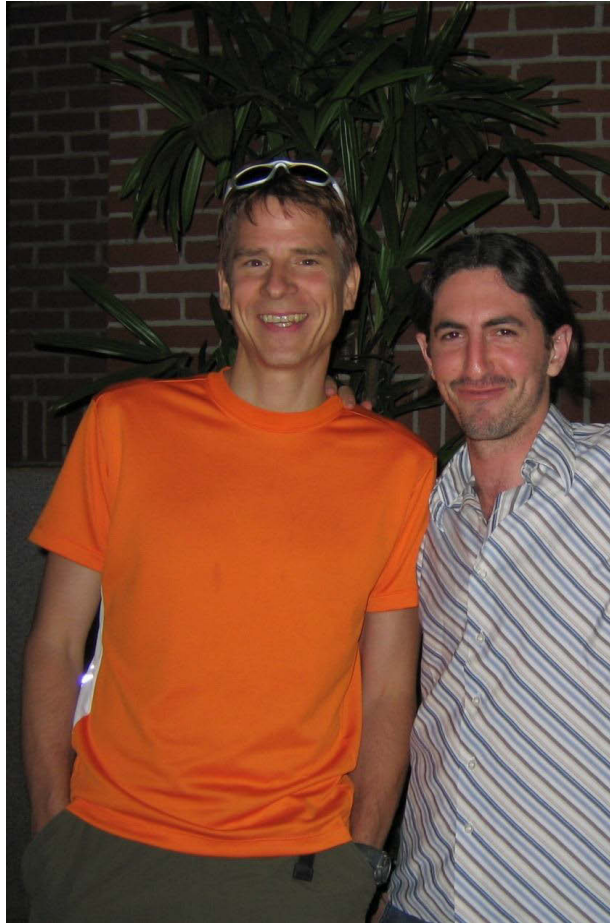
truth. My mother never said “her son is a doctor, but not the kind that helps people” – on the contrary – my parents always made me feel that the choice I had made was making them proud and happy, and it, therefore, made me feel the same, constantly. Yarin my brother kept me sane through hard times and chatted with me daily about research and about the world, keeping me focused and reminding me what’s actually important in my work. His focus and clarity were a guideline in my work. I wish I had his character for those 4 years. Had I had it, I would have been done much faster, and probably done an even better job. He is an idol of mine. As for Kelsey and Ginger – in the words of Randy Pausch: “I am good. But not good enough to talk about them”, and as such their contribution is something I keep for myself. The audience doesn’t have to get all the information, as once said by Brecht.

Finally, no words can describe the amount of gratitude and admiration I have for my advisor, mentor, friend, and father for the last 4 years. Christof is one of the people that make me want to be a better person, and forever shaped every thought and action I will ever make. Any achievement I have in my life, from now on, should grant Christof a residual. Our talks into the nights, work through the day, collaboration in science and beyond are worth more to me than anything else I studied in all my years in academia. I could have never done what I did without him. From our first encounter in a small room in Jerusalem, at a time where I was not at all confident about my ability to leave my family behind and make this leap, hearing his description of the relationship between an advisor and a student – using the words “Doctor father” inspired me to think that I would not lose family by leaving Israel for Caltech, but actually gain an additional family member. And indeed this was the case. Christof suggested I go to Burning Man my first summer as a graduate student because of its theme – consciousness. Convinced me to film this experience which completely turned my world

upside-down. Christof allowed me to take time and go to an art conference (Siggraph) in San Diego to “get ideas”, and knew before I did the benefits of open mindedness in science. Christof made me feel that my background in breaking codes is useful in understanding the ultimate black-box of codes - the brain - and thus made my entire background and past life before starting to work for him make sense and seem like it was all part of a grand plan that only he knew of.

Christof made me laugh in hard times, put me in my place when I broke loose, told me which book to read or which movie to watch at the right times, knew which buttons to press to make me do my best, and squeezed excellent research out of me that I did not know I had. He exposed the scientist in me, which I did not even know was there.

I feel that my time under Christof's umbrella made me not only a better scientist, but hopefully a better person.



June 2005

Abstract

At any given moment, our brains are bombarded with enormous amounts of information from the environment. External stimuli from the senses travel through our eyes, nose, or skin, and internal reflections and imagery travel from within. All these stimuli are processed in parallel and compete with each other toward one ultimate goal: becoming the single percept which we are aware of at this unique present moment.

This work studies the way by which this competition occurs in our brains. The mechanisms and the methods which allow the brain to overcome this load of information by selecting one of many thoughts to reach the upper level of our consciousness.

We studied healthy controls in eye-tracking experiments where they viewed sets of images in different tasks. In the first task, subjects freely viewed images with semantic high-level cues such as faces and social scenes. Subjects showed a significant tendency to rapidly attend to the faces in their first fixation. In a second task, subjects viewed similar images in a search task where they were instructed to find objects or faces in the scene. Subjects showed the same tendency to look at faces independent of the task. In a following study, subjects were looking at images with text and phones as control and were shown to look at text rapidly and early, but not as much as faces. Phones, as control, showed no increased rapid attentional attraction. In order to test the magnitude of the effect, we had subjects participate in a third experiment where they were instructed to refrain from looking at these objects, but were not able to do so as easily for faces as they were for text and phones. This suggests an innate mechanism that draws our

attention to faces and social scenes early — even in situations where other stimuli compete for our attention. Faces win the competition for our attention regardless of the task. We used this striking result to modify an existing computer model for bottom-up attentional allocation to better predict human's fixation in images.

The results were tested additionally in two groups of individuals with disorders that manifest themselves primarily in decreased social attention: autism and agenesis of the corpus callosum (AgCC; subjects who are missing the bridge between the left and right hemispheres in the brain). Individuals with these disorders were tested in the same paradigms and indeed showed decreased attention allocation to faces or social cues in the images. AgCC subjects showed an even lower level of interest in social cues. While the results suggested a competition for attention that has social cues win over alternative cues in healthy controls, the psychiatric disorder groups show no such effect. In order to test the competition in the brain even further we tested individuals with epilepsy undergoing brain surgery, who were implanted with electrodes to record from single neurons in their medial temporal lobe (MTL). These patients participated in a task where they were projecting their thoughts of one of various concepts onto a computer screen, in real-time, using a decoder that interpreted their thoughts and imagery. Patients performed a task in which they were instructed to think of one of four concepts, and by accurately doing so were fading into an image representation of that concept on the computer screen. Multiple images that were shown on the screen simultaneously while the patient tried to suppress one and maintain an imagery of the other allowed us to directly create a situation where competition between various external stimuli, and in turn multiple brain regions, is tested. Subjects were able to reach high level of control of their single MTL neurons after very little training. In this direct measure of the competition between brain regions and

neurons in the brain, we show that attention can be modulated to direct the flow of information to one or the other area, even though the external stimuli from the environment is identical. We used the results from the fading experiment to construct a model of the mechanism by which competition between external stimuli and internal imagery modulates attention in the brain.

Subjects who were able to reach an even higher threshold of control of a single neuron played a computer game in which they were controlling an airplane on the screen using their thoughts alone. These subjects reached a high level of control suggesting an ability to use single neurons in the MTL for brain-machine interfaces with very high accuracy.

Finally, we report various case studies from experiments involving direct measures of attentional allocation by individuals with face blindness (Prosopagnosia) who performed poorly in tasks involving competition between facial and social attention attractors; a subject with no amygdalae who was unable to direct her attention to fearful entities including images of herself posing while displaying fearful emotions; identical twins who showed an extremely high correlation in their attentional allocation metrics – both eye-tracking and individual rating of interest in images; and a similar high correlation between a mother and her autistic son.

Altogether, these results shed light on the processes and the mechanisms undergoing in our brain, in the milliseconds between the moment information starts flowing in our brain from the environment, through the spotlight of attention that selects which of various inputs will be selected to reach our consciousness.

Table of Contents

<i>Acknowledgment</i>	<i>ix</i>
Collaborators	ix
Friends	xiv
In loco parentis.....	xvii
People	xvii
<i>Abstract</i>	<i>xxvii</i>
<i>List of Figures</i>	<i>xliii</i>
<i>List of Tables</i>	<i>li</i>
<i>Nomenclature</i>	<i>liii</i>
Anatomical abbreviations.....	lv
<i>1. Foreword by Yoram Kaniuk</i>	<i>1</i>
<i>2. Prelude</i>	<i>5</i>
The application.....	9
Consciousness.....	10
ICNC	11
Marathon	13
caL01jan	15
Kelrof.....	16

Dreams from my Father	16
Kelsey	17
<i>3. Introduction: Setting the Stage.....</i>	<i>19</i>
<i>4. A question and a theory.....</i>	<i>25</i>
<i>5. Subjects.....</i>	<i>31</i>
Autism	33
Neuropsychology.....	34
Diagnosis	36
ADOS	36
ADI-R.....	37
Benton.....	39
STAI	39
WAIS.....	39
Relevance to attention study	40
AgCC.....	42
Competition.....	42
Background	43
Behavior	44
Are these “split-brain” patients?	45
Relevance to attention	46
Amygdala lesion.....	48
Relevance to attention	50
Prosopagnosia.....	51
Blink	51

Cats	52
Law and Order	56
Jay Leno.....	58
Face blindness	59
Background	61
Relation to attention.....	62
Epilepsy	64
Background	64
Surgery.....	66
Surgical procedures.....	67
Electrode placement.....	69
<i>6. Attention and Competition.....</i>	<i>75</i>
Introduction.....	76
There can be only one	76
Attention	79
Competition.....	80
Behavioral data in neuroscience	81
Neural basis for competition	82
Models of attention as competition between brain resources on the conscious percept	87
Multiple stimuli compete for neural representation in visual cortex	87
Competition can be biased by top-down and bottom-up mechanisms	88
<i>7. About Face</i>	<i>93</i>
Introduction.....	97
Prior studies about faces	100

1. Humans can recognize faces in extremely low-resolution images	105
2. Humans can recognize faces in extremely low-resolution images	106
3. Facial features are processed holistically	107
4. Both internal and external facial cues are important and they exhibit nonlinear interactions	108
5. The configural relationships is independent across width and height.....	110
6. Vertical inversion dramatically reduces recognition performance	111
7. Contrast polarity inversion dramatically impairs recognition	112
8. The visual system starts with a rudimentary preference for face-like patterns.....	113
9. The human visual system appears to devote specialized neural resources for face perception.....	114
10. Latency of responses to faces in IT (120 ms) suggest a feed-forward computation.....	115
Face detection.....	117
8. Methods	127
Psychophysics metrics.....	128
Eye tracking	129
Device	129
Images.....	133
Presentation times	139
Exposure	140
Eye tracking	143
Modeling using the saliency model	144
Single-neuron recordings	145
Spike detection.....	145
LFP Analysis.....	146
60 Hz noise removal	146
Low-pass filtering	146

Experiment presentation.....	147
Internal Review Board and HIPAA.....	149
<i>9. Observers Are Consistent When Rating Image Conspicuity.....</i>	<i>151</i>
Art.....	153
The museum stroll.....	153
Experiment.....	154
Scale.....	155
Outcome.....	156
Introduction.....	157
Methods.....	159
Stimuli.....	159
Participants.....	160
Presentation.....	161
Conspicuity rating.....	161
Experiment 1.....	162
Experiment 2 – Control for the relevance of memory.....	167
Experiment 3 – Effect of presentation duration.....	167
Results.....	168
Experiment 1 – Phase 1 – Conspicuity rating.....	168
Intra-observer correlation.....	168
Inter-observer correlation.....	173
Experiment 1 – Phase 2 – Memory.....	175
Experiment 2 – Long-term longitudinal consistency.....	178
Experiment 3 – Effect of presentation duration.....	180
Discussion.....	181

Shirt	183
<i>10. Attention and High-Level Cues. Part I – Faces</i>	<i>185</i>
Methods	187
Experimental procedures	187
Results	190
Psychophysical results	190
Discussion	196
<i>11. Attention and High-Level Cues. Part II – Text</i>	<i>197</i>
Introduction	199
Methods	200
Subjects	200
Stimuli	200
Experimental paradigm	201
Results	204
<i>12. Attention and High-Level Cues. Part III – Faces and Text</i>	<i>213</i>
Methods	217
Experimental procedures	217
Images	220
Analysis Metrics	222
Fixations	222
Saccades	222
Baseline calculation	224
Results	227

Psychophysical results	227
Experiment 1 (“free-viewing”)	227
Experiment 2 (“control for the relative effect of size”)	229
Experiment 3 (“search”)	230
Experiment 4 - control for the effect of adaptation.....	234
Discussion.....	237
<i>13. What happens in your brain before a saccade is initiated.....</i>	<i>245</i>
Introduction.....	247
Methods.....	251
Results.....	255
Free viewing.....	255
Early Saccades	255
Inter-saccadic changes in LATER units.....	259
LATERian Race-to-Threshold competition.....	262
Top-Down Influences	265
Discussion.....	268
Bottom-up Saliency	268
Top-Down Influences	271
Early Saccades	272
Latencies in the visual system.....	273
<i>14. Predicting Human Gaze Using Low-Level Saliency Combined with Face Detection.....</i>	<i>281</i>
Methods.....	284
Combined face detection with various saliency algorithms	284
Results.....	288

Assessing the saliency map models	288
Discussion.....	292
<i>15. Computer Model of Faces and Text Gaze Attraction.....</i>	<i>293</i>
Methods.....	295
Results.....	299
Model analysis	299
Discussion.....	301
<i>16. Decoding What People See from Where They Look</i>	<i>305</i>
Methods.....	308
Experimental setup	308
Decoding metric.....	310
Results.....	314
Discussion.....	321
<i>17. Visual Attention Allocation in Individual with Autism</i>	<i>325</i>
Introduction.....	327
Methods.....	330
Results.....	334
Discussion.....	342
<i>18. Anecdotal cases</i>	<i>345</i>
Monozygotic (identical) autistic twins.....	346

Introduction.....	346
Method	347
Results.....	348
Correlation.....	348
Categories.....	349
Correlation without sequences.....	350
Perfect agreement.....	350
Clustering.....	351
IAPS.....	351
Eye-Tracking.....	352
Conclusions	355
Prosopagnosia.....	356
Results.....	357
Discussion.....	364
Amygdala lesion.....	366
Experiment I – Attention allocation to faces.....	366
Methods	371
Results	373
Experiment II – Attention allocation to faces carrying emotions.....	375
‘Discussion’	376
Results.....	380
Discussion.....	385
Parent of autism subject.....	387
Results.....	389
Conclusion.....	391
Agensis of the Corpus Callosum.....	393

Methods..... 394

Results..... 395

Discussion..... 399

19. Attention in Marketing 401

 The construct of attention in advertising context 403

 The concept of attention..... 404

 Measurement of attention..... 407

 Computational neuroscience of visual attention 408

 The model of bottom-up attention and saliency of Itti, Koch, and Niebur..... 410

 Potential contributions of computational neuroscience of visual attention to marketing research..... 413

 Application of computational modeling of visual attention to evaluation of print advertisements 414

 Conclusions 420

 Directions for future research..... 420

20. Visual Projection of Thoughts from Single Neurons in Humans 423

Introduction..... 425

Methods..... 428

 Experimental paradigm..... 434

 Screening..... 434

 Mini screening 435

 Fading..... 435

 Full procedure 437

 Decoding..... 438

 Setup 441

 Sites 443

 Multi-units versus single-units..... 447

 The purpose of the fake trials 447

Bootstrap testing of statistical significance for task performance	448
Image saliency.....	449
Results.....	452
Invariant representation of a concept.....	466
Comparison of real/fake feedback	468
LFP analysis of the data	468
Difference between the different screenings	473
Behavior	481
Discussion.....	483
Feedback in the vision system.....	484
<i>21. Modeling Attention</i>	<i>489</i>
Method	491
Results.....	497
Discussion.....	500
<i>22. Computer Games</i>	<i>503</i>
Introduction.....	506
Methods.....	509
Experiment	509
The game	509
Configuration.....	511
Results.....	516
Discussion.....	525

<i>23. Future</i>	529
21 questions	530
Predictions	533
<i>24. Conclusion</i>	537
Summary.....	539
Contribution and significance of this work.....	545
Consciousness	547
Final sentence	548
<i>Index</i>	549
<i>Appendix I. Statistics</i>	551
<i>Appendix II. Getting the Data</i>	553
<i>Appendix III. Interview with SM</i>	561
<i>Appendix IV. Brain for Sale</i>	575
<i>Appendix V. Patent</i>	577
Ingredients	577
Suggested applications	577
<i>References</i>	581

List of Figures

Figure 1 - Subject SM's CT scan	49
Figure 2 - Pictures of two people who seem identical to a subject with Prosopagnosia.....	52
Figure 3 - Illustrations of the faces as seen by people with Prosopagnosia	55
Figure 4 - District Attorney Jack McCoy and Detective Kevin Bernard	57
Figure 5 - Caricatures of Jay Leno and Tom Cruise	59
Figure 6 - A photograph showing one of the patients	70
Figure 7 - Electrodes implanted in patient's brain	71
Figure 8 - MRI image of one of the patients	72
Figure 9 - Regions we recorded from in the medial temporal lobe.....	74
Figure 10 - Processing pathways in the visual stream	83
Figure 11 - People see faces everywhere.	95
Figure 12 - Example of a face pop-out effect.....	98
Figure 13 - Subjects are able to recognize familiar faces in low resolution	105
Figure 14 - Images with only contour information difficult to recognize	106
Figure 15 - Holistic versus feature-based processing.....	107
Figure 16 - Identity recognition based on purely external cues	109
Figure 17 - Changes in configural relationship in faces do not hurt perception.	110
Figure 18 - Inverted faces: The Thatcher illusion.....	111
Figure 19 - Negative contrast image from a Beatles' album.....	112
Figure 20 - Newborns preferentially orient their gaze to the face-like patterns.....	113

Figure 21 - An example of the FFA in one subject	114
Figure 22 - Examples of an IT cell's responses to variations on a face stimulus	115
Figure 23 - Feature types used by the Viola and Jones algorithm	118
Figure 24 - Features selected by the AdaBoost	120
Figure 25 - Cascade architecture of the Viola and Jones algorithm.....	121
Figure 26 - Examples of the results of the Viola and Jones algorithm	123
Figure 27 - Examples of the Viola and Jones results on complex images.....	124
Figure 28 - Example of different eye saccades due to different task	130
Figure 29 - Eye tracker	131
Figure 30 - Fixations accuracy in the eye-tracker.....	132
Figure 31 - Illustration of the vision experiment: block 1. Free viewing	135
Figure 32 - Illustration of the vision experiment: block 2. Search task	136
Figure 33 - Illustration of the vision experiment: block 3. Free viewing	137
Figure 34 - Illustration of the vision experiment: block 4. Memory block.....	138
Figure 35 - Images distribution across tasks	139
Figure 36 - The exposure normalization process.....	143
Figure 37 - Image classes used in the study	160
Figure 38 - Conspicuity rating experiment design	165
Figure 39 - Observer 4's responses to fractal images	168
Figure 40 - Intra-observer correlation.....	170
Figure 41 - Observers become more consistent after repeated exposure	172
Figure 42 - Inter-observer correlation.....	174
Figure 43 - Recognition memory	176

Figure 44 - The saliency experiment t-shirt worn by our subjects	183
Figure 45 - The saliency experiment t-shirt worn by our experimental team	184
Figure 46 - Examples of stimuli during the “free-viewing” phase	189
Figure 47 - Extent of fixation on face regions-of-interest (ROIs) during “free-viewing”	192
Figure 48 - Number of faces visited for subject 2, experiment 1	193
Figure 49 - Percentage of faces visited	194
Figure 50 - Number of fixation in faces for subject 1, experiment 1	195
Figure 51 - Examples of an image transformed used in this experiment	202
Figure 52 - Images with subject’s scan paths overlaid	202
Figure 53 - Comparison of fraction of trials with fixations in the region of interest	206
Figure 54 - Comparison of model performance	209
Figure 55 - Relative weight of text channel versus	210
Figure 56 - Examples of images from the three categories: Faces, text, and cell phones	219
Figure 57 - Illustration of the computation of the saccade planning time	223
Figure 58 - Illustration of the computation of the baseline	225
Figure 59 - Extent of first fixation on ROI during free-viewing task.....	228
Figure 60 - Saccade planning durations and proportions of fixations in avoid/search task ...	232
Figure 61 - Examples of saccadic latency fits of two subjects	253
Figure 62 - Percentage of early saccades landing on a face region of interest	256
Figure 63 - The proportion of fixations landing on faces/text	258
Figure 64 - Fixations to the faces reflect increased information and mean rate of rise	261
Figure 65 - Competition and mean rate of rise affect the saliency histogram	264
Figure 66 - Proportions of fixations on the face in the search task.....	266

Figure 67 - Differences in relative saliency affect fixation location.....	270
Figure 68 - Latencies in the visual pathway	274
Figure 69 - Modified saliency model	286
Figure 70 - Comparison of the area-under-the-curve (AUC) for an image	289
Figure 71 - Comparisons of the SM, SM + VJ and GBSM, GBSM + VJ	291
Figure 72 - Illustration of the combined saliency map with a semantic channel added	296
Figure 73 - Performance comparison for all 231 images.....	300
Figure 74 - Comparison of the Ideal AUC normalization and the saliency models.....	303
Figure 75 - Examples of scanpath/stimuli used in the experiment.....	309
Figure 76 - Illustration of the AUC calculation	313
Figure 77 - Decoding performance with respect to image pool size	315
Figure 78 - Performance of the 9 individual subjects.....	318
Figure 79 - Decoding performance based on feature maps used.....	320
Figure 80 - Examples of stimuli used.....	332
Figure 81 - Proportion of first fixations in autism	335
Figure 82 - Proportions of first fixation in the region of interest	341
Figure 83 - Rating of images for the identical autistic twins	348
Figure 84 - Rating of the entire dataset for the identical autism sisters.....	349
Figure 85 - Distributions of answers	352
Figure 86 - Examples of scanpath for the autism identical twins	353
Figure 87 - Examples of scanpath for the autism identical twins	353
Figure 88 - Examples of scanpath for the autism identical twins	354
Figure 89 - Proportion of first fixations on a face in “free viewing” in prosopagnosia.....	358

Figure 90 - Proportion of first fixations on a face in “search task” in prosopagnosia	360
Figure 91 - Failure to correctly identify the face in Prosopagnosia.....	362
Figure 92 - SM looking at faces in blocks I and II	367
Figure 93 - Face perception and attention systems	369
Figure 94 - Scanpaths of subject SM in natural scenes images with people	372
Figure 95 - Scanpaths of subject SM in full-sized faces images.....	372
Figure 96 - Subject SM looks mainly at the mouth	374
Figure 97 - Fearful faces from Google image search.....	379
Figure 98 - SM judgment of emotions in images.....	381
Figure 99 - SM judgment of emotions in images in a task guiding her to fixate the eyes	382
Figure 100 - Emotions performance change between the task.....	383
Figure 101 - Numerical ratings of images of mother and her autistic son.....	390
Figure 102 - Fraction of first fixation in face region for AgCC subjects in free-viewing.....	395
Figure 103 - AgCC subjects’ proportion of first fixation on face region.....	396
Figure 104 - Distribution of first fixation on the face for AgCC subjects in search task.....	397
Figure 105 - Proportion of first fixation in face regions for the search task in AgCC.....	398
Figure 106 - Visual processing pathways.....	409
Figure 107 - Computational model architecture	411
Figure 108 - Bottom-up attention to ad 1	416
Figure 109 - Bottom-up attention to ad 2	418
Figure 110 - Example of a single session from a patient.....	429
Figure 111 - Example of a single session from patient 1.....	431
Figure 112 - Example of a single session	432

Figure 113 - Example of a single trial	433
Figure 114 - Illustration of the experiment.....	436
Figure 115 - Illustration of the decoder	440
Figure 116 - Illustration of the experimental setup	442
Figure 117 - Units distribution	444
Figure 118 - Illustration of the bootstrapping technique.....	450
Figure 119 - Success rate	458
Figure 120 - Learning for a single patient	460
Figure 121 - Additional example of a learning effect	461
Figure 122 - Group “learning”	462
Figure 123 - Neuronal activity dependence on imagery/vision.....	464
Figure 124 - Mental control of the feedback	465
Figure 125 - Responses of channel 45, left anterior hippocampus, in patient 399.....	467
Figure 126 - LFP traces from patient 405, channel 53.....	470
Figure 127 - LFP traces from a single patient.....	471
Figure 128 - Power spectrum density estimation.....	474
Figure 129 - Power spectrum density estimation from channel 53	475
Figure 130 - One-way frequency by frequency analysis of variance for channel 3.....	476
Figure 131 - One-way frequency analysis of variance for channel 53.....	477
Figure 132 - ANOVA between different frequencies in the "during" interval.....	478
Figure 133 - STFT averaged across image repetitions	479
Figure 134 - Mean LFP traces and \pm S.E.M.....	480
Figure 135 - Map of the visual system in humans	485

Figure 136 - An illustration of the computer model	495
Figure 137 - Estimation of the attention level for each patient	499
Figure 138 - Illustration of the configuration used in the experiment.....	512
Figure 139 - Illustration of the computer game played in the experiment.....	514
Figure 140 - Illustration of the game played and the single neuron eliciting the activity	518
Figure 141 - Distribution presentation of the performance for the 2 patients	521
Figure 142 - Performance of neighboring neurons in the game	523
Figure 143 - Performance of neighboring neurons in the game	523
Figure 144 - Provisional patent application	580

List of Tables

Table 1 - Summary of methods and processes for faces detection	104
Table 2 - Eyelink-1000 parameters.....	133
Table 3 - Order of categories used in experiment 1	163
Table 4 - Order of categories used in experiment 2	179
Table 5 - Summary of timing and accuracy results for the 4 experiments	236
Table 6 - Studies linking timing and interaction between brain regions and attention	277
Table 7 - Information about autism subjects	333
Table 8 - Details on AgCC subjects	394
Table 9 - Competition across brain regions	446
Table 10 - Parameters and variables used in the fading attention model.....	494

Nomenclature

2 AFC. Two-Alternative Forced Choice. The term is used in several psychophysical experiments where the subject has to respond by making a forced decision between two possible alternatives.

EEG. ElectroEncephaloGraphy. Throughout the text we use the term EEG to refer to the electrical potentials recorded typically on the scalp, outside the skull, but not intracranially. This technique was introduced by Berger in 1929 (Berger, 1929).

BOLD. Blood-Oxygen-Level-Dependent

CT. Computerized Tomography

EPSP. Excitatory PostSynaptic Potential

ERP. Event-Related Potential

fMRI. Functional Magnetic Resonance Imaging

IPSP. Inhibitory PostSynaptic Potential

IQ. Intelligence Quotient

ISI. InterSpike Interval. Time difference between successive action potentials

LFP. Local Field Potential

LTP. Long-Term Potentiation. (Bliss & Lomo, 1973; Kandel, Schwartz, & Jessell, 2000)

MRI. Magnetic Resonance Imaging

MEG. Magneto-EncephaloGraphy

MUA. Multi-Unit Activity

PCA. Principal Component Analysis

PET. Positron Emission Tomography

r². Correlation coefficient

RT. Reaction Time

ROC. Receiver Operator Characteristic

SNR. Signal-to-Noise Ratio

SVM. Support Vector Machine

SUA. Single-Unit Activity

SD. Standard Deviation

SEM. Standard Error of the Mean

TMS. Transcranial Magnetic Stimulation (Kamitani & Shimojo, 1999; Pascual-Leone et al., 1998; Ruohonen, 1998)

Anatomical abbreviations

ACC. Anterior Cingulate Cortex

Amy. Amygdala. Almond-shaped mass of gray matter in the anterior portion of the temporal lobe. Also called *amygdaloid nucleus*. (Latin, almond)

EC. Entorhinal Cortex (from Greek entos, within.) Structure within the rhinal cortex. Brodmann's area 28

FFA. Fusiform Face Area

FEF. Frontal Eye Fields

Hip. Hippocampus. A ridge in the floor of each lateral ventricle of the brain that consists mainly of gray matter. (Late Latin, a sea horse with a horse's forelegs and a dolphin's tail (from its shape in cross section), from Greek hippokampos : hippos, horse. + kampos, sea monster)

IT. Inferior Temporal cortex

LGN. Lateral Geniculate Nuclei

LIP. Lateral Intraparietal Area

MTL. Medial Temporal Lobe

mPFC. Medial Prefrontal Cortex

PHC. Parahippocampal Cortex

PPA. ParaHippocampal Place area

V1. Primary visual cortex, Brodman area 17



Foreword by Yoram Kaniuk

Every time a friend succeeds, I die a little.

Gore Vidal

H²ow does our brain give rise to our conscious mind? I know very few people who would dare to tackle this question as a Ph.D thesis. Consciousness - a phenomenon that seems so elusive and so out of our reach, somewhere between science-fictions movies, and Jules Verne's novels. A topic that poets and sonnet writers delve upon so regularly, and yet that remains elusive to us. This is the topic of Moran's work.

If I were to trust one person to be able to formulate the question correctly and offer a path towards addressing this topic, it would be Moran. This is not only because of the clarity of his mind and his intuitive ability to see the world purely and simply, but because of his ability to see the story in it. To see the threads of reasoning that connect bits and pieces of information to an aesthetic view of the world. While Moran is a brilliant scientist and in my experience is capable of excelling in everything he touches, the characteristic that impresses me the most about Moran is his ability to make everything he does interesting. Purely by being curious about it, and wanting to know and understand it.

Moran's scientific abilities far exceed my own, and much of his thesis was above my head. Nevertheless, I read Moran's thesis cover to cover, drinking in the bits of wisdom and witty comments within the pages of this marvelous work. Moran's thesis is a unique combination of a thorough investigation and a wonderful story, written like a truly great book. The story of our brains, our selves, told from various angles and perspectives. A personal tale of a young scientist exploring the world both within and outside himself. The story of disorders of the brain, as told by the scientists who study them, the individuals afflicted by them, and by the people who love those individuals. Throughout this work, Moran never fails to provide us with exciting examples and glimpses into the lives, difficulties, and adventures of these fascinating individuals. Moran provides us with a deep look into some of the rarest psychological disorders, and in so doing helps to build an entirely different view of our world. Such are the marvelous stories about the face-blind subjects, or the woman lacking the region of the brain which allows us to experience fear.

² Translated from Hebrew by the author.

This work describes how a series of mechanisms in our brains compete and vote to decide what will penetrate our attention, our very awareness – what we will know and experience. In a marvelous work that includes the results of experiments recording directly from the exposed brains of patients undergoing brain surgery, Moran takes a deep look inside the brain and uncovers the mystery of our perceptions.

I recommend everyone to read this thesis, not just as the sum of multiple research studies covering years of rigorous study, but also as a book. Read it as a description of a fascinating facet of our lives, both from the outside and from within our own brains. I think that as such, it offers a remarkable view of who we are, why we are, and what drives our attention, intention and minds.

I have known Moran for over fifteen years now. I have known him as a student, a caregiver, a companion, an actor, a singer, a writer, and as a friend. Many things have changed in both of our lives since we first met in the halls of a high-school in Tel-Aviv. There we met when I started a writing workshop where I led high school students in writing short stories. Moran and I shared stories and adventures in this workshop for two years. And kept exchanging manuscript ever since. For nearly thirteen years now. Moran and I met recently in California where we both attended the premiere of a movie based on a book of mine. We traveled in a convertible to places we both experienced in America – 60 years apart. That's when he asked me to write the foreword for his thesis that I then learned was dedicated to me. Israelis usually dedicate things to the dead, honoring their lives. In the U.S. people honor you while you're still alive. When I met Moran last I was somewhere in between, just recovering from a near-death experience which was the subject of a book of mine that came out at the time. I am often times implicated in a genre of writing known as a stream of consciousness. My experience with Moran, and the fact that devoted his graduate studies to investigating that consciousness, make me be proud to be part of that genre. The fifteen years of knowing Moran brought many adventures to us two, many experiences and various conversations and interactions. Many times in my life I dreamt of being

a scientist. But through the eyes of the grandson I never had I get to vicariously live the life of a scientist. I am thankful to Moran for including me in his life endeavors and for choosing me as a friend and a companion in his life journey. I am confident that the story of Moran has yet to bring many more stories and answer to many others, and I only wish to be here, conscious, long enough to be a witness to them all.

Yoram Kaniuk

Tel-Aviv, 2009



Prelude

To: 'Moran'
Cc:
Subject: Earn a University Degree based on your professional experience.

Want the degree but can't find the time?

WHAT A GREAT IDEA!

We provide a concept that will allow anyone with sufficient work experience to obtain a fully verifiable University Degree.
Bachelors, Masters or even a Doctorate.

Think of it, within four to six weeks, you too could be a college graduate.

Many people share the same frustration, they are all doing the work of the person that has the degree and the person that has the degree is getting all the money.

Don't you think that it is time you were paid fair compensation for the level of work you are already doing?

This is your chance to finally make the right move and receive your due benefits.

If you are like most people, you are more than qualified with your experience, but are lacking that prestigious piece of paper known as a diploma that is often the passport to success.

CALL US TODAY AND GIVE YOUR WORK EXPERIENCE THE CHANCE TO EARN YOU THE HIGHER COMPENSATION YOU DESERVE!

+1(630)225-5047



fter 4 years of research at Caltech, which is by far the best place I have spent 4 years in my entire life, and after 4 years which are likely candidates to have been the best years of my life, I am reaching the end of one particular road. A small file, approximately 900kb in size, a font, a style, a couple of funny or witty quotes in the beginning of each chapter, a theme, a few spelling/grammar mistakes and some typos, a couple of brilliant ideas hidden within repetitions of things said by others in nicer words. My thesis is complete.

After 4 years of research, for the first time in my life I am not tempted by junk emails cluttering my mailbox offering me a Ph.D. within 3 days at the University of Phoenix for the small amount of \$1499. I know now that my chances of getting a Ph.D. from Caltech, for a much cheaper price are higher, and could well happen within a shorter time. The end of a road that started – not 4 years ago when I first set foot at the doors of Christof’s office, coming for a rotation in his lab, only to realize from his secretary Shannon that he had forgotten about my arrival – but a journey that started 10 years ago. When I was turning 21.

It was a rainy winter day in Tel Aviv. I was grounded to my military base for having too long hair 3 times in a row. Stood before a military judge and pleaded guilty. Was convicted and sentenced to 10 days of being grounded to the military base. It wasn’t my first trial in the army, and would not be my last before I ended my army duty. I had 10 more months until I finished my service. By then I would be 21. This looked like a very serious age. A grown up. In Israel you are eligible to run for office in the parliament by that age. When I was 4 years old I got this Bazooka Joe chewing gum that had a fortune in it – common in Israel. It told me then that by the age of 21 I would reach the moon. I still had 10 more months to do so....

It was around January of that year and I was sweeping the floor of the dusty and dirty room where I was to spend the next 9 days of semi-punishment, when I got a phone call from my father. I knew he wouldn't be happy with my telling him about the results of the trial. Yet, again, his son was convicted for rebelling against the army. I think he liked it on one hand, but hoped for a calmer, less "exciting" army service – or for that matter, life – from his older son. I picked up the phone and was surprised to learn that the subject was in fact far from what I had imagined. "Tomorrow is the last day to apply to the university if you want to start studying next year", my dad said. I was to end my army duty on the 22nd of October 1998. The school year in Israel was to begin October 25th. 3 days later. A long weekend. In my wildest nightmares had I not imagined I would go to the university straight from the army. In Israel it's common to take some time off after being reduced to a military number for over 3 years and told when to bathe, sleep, or pee day and night. A 3 month vacation in India, backpacking in Guatemala, trekking coast to coast in the U.S., all seemed a lot more appealing to me at the time. I was quick to tell my father that I had no intentions of applying to the university straight after the army. 3 days after it, to be exact. We had a vivid over-the-phone debate. My father pressed the idea that I should go there, or at least have a clear idea what it was that I wanted to "do with my life when I grew up". "If you think that backpacking will make you happy, I will support you in doing so. But as long as you're not SURE what it is that you want to do – I suggest applying to the university. Just apply. If you realize later on that you don't want that – you can always not go". I didn't have a good counter argument at the time. Mostly - I had 9 days to do something with myself, and spending them prepping for the Israeli-SATs didn't seem like a terrible idea....

In Israel, applying to the university consists of 2 components: your high-school grades, and an SAT-like test titled the “Psychometric exam” that grants you a score ranging from 200-800. The two components are combined in a complex formula to give a number. The numbers are ranked and are the only thing that is used to determine if you will or will not be accepted to a program in the university. If many people in that year are eager to study computer science and there are only 150 seats in that year, the 150 people that applied for CS with the highest score will be accepted – the rest will have to take some other major. Yes, in Israel one picks his major right away – before starting college. The Bachelor degree takes 3 years and is very intense and focused. From day one you take classes in the field you chose as your major. Almost no elective classes, and very condensed program. You can’t choose your classes, or their order, or time. You basically continue your army service for 3 more years, only this time it’s called “first degree”. Similar to the doing time for murder, only with no parole after a third of the time for good behavior.

My dad showed up that evening with the application book. A 251 page book with the list of all possible majors and details about the likelihood of getting accepted into each as well as description of all the classes and programs entailed in each major. I was fascinated. So many exciting things that I could thoroughly learn. Yet... such a committed decision. Based on a short 5 page description I should choose a career. A path. Something that will determine my next 3 years possibly. How could I do that?

The application

I saved my application letter from January 11th 1997. Written in a very old version of Microsoft Word, on a laptop the size of a huge modern server. Saved to a 5¼” floppy disk, and submitted 2 hours after the deadline (a phenomenon that would stay with me for years later), under the “rule of 80” (see box on right; later – during my work with Chrisotf, I learned from him a more concise wording for the rule, which he sometimes used: “Don’t let the perfect be the enemy of the good”). The application letter. Based partially on Albert Einstein’s immortal words: “I want to know God’s thoughts – the rest are details”, I ambitiously set forth my beliefs and non-beliefs in academia – based on absolutely no knowledge of the system or the ways by which it works. In an arrogant fashion I started by explaining what university was for me, and what I thought could be improved in it – not having set foot in one ever before – and went on to describe my life and interests, only to end with the “bottom line”:

The 80 percent rule

The 80 percent rule originally came about when I was in my first year of studies. I took dozens of courses back then... many of which I had to submit an essay at the end of. I obviously started working on those around midnight the night before. Given that timeframe I had no chance of doing the good job I was willing to do, but I had to do something.

At first I had those common conflicts: maybe I should cancel the class... afterall I came to learn and not to just pass the exam... maybe it is better to wait for next year... I will have a lot of time then, and I will do a better job... maybe I should postpone the submission, and do it thoroughly...

Obviously these were all lies. I would not make a greater effort in the future. Everything would be just the same... That’s when I thought about the 80 percent law. A law that came to remind me that I came to take classes and learn and should spend most of my time doing reading and learning rather than writing and providing replicas of words that were already said by others.

A grade of 80 is better for that. It saves time, but still forces you to work. That’s that. I have no need for a better grade. I will write whatever I can as long as it meets the minimal standards. No intelligent remarks, no deep reflections. Nothing. The minimum... it will get me 80, and I will move on. Fine by me.

The rule of 80 says that if you are a person that looks for superficial vast amount of knowledge rather than specialization then then you can allow yourself to aim towards the 80 in whatever you do, as long as you get an 80 in

I felt that science deals with many questions – which I was sure are interesting to people and are important and beneficial to the world. But as far as I was concerned there are only 3 questions that I though SHOULD be interesting to a science that I would be part of, and I

would let YOU accept me into your program if I would be allowed to study one of these 3.

And then I set forth to list the 3 questions that I felt really were important for science:

Understanding the beginning of life.

Figuring out whether there are alien life forms outside of Earth.

Understanding the way by which consciousness works.

... the rest were details.

Consciousness

Consciousness seemed like the most tangible scientific question to me. An uncharted land with a sense of everything. Studying consciousness would allow me to be a physicist, a physician, a biologist, a psychologist, computer scientist. Everything. I didn't need to make a choice.

Two years into college I focused on Physics as my major for my bachelors and ended up getting a B.Sc in Physics. I wasn't an excellent student, but I was definitely faster and more efficient, having finished my studies a year earlier than my peers. The remaining two years were dedicated to going in a different direction, studying the "Philosophy of Science", focusing my Masters research question on the "ability of science to study notions such as consciousness", titling my Master's thesis: "On the mind-body problem. Without the mind", and claiming that humans are merely bodies, and that consciousness and all the phenomenological effects that are associated with it are purely neuronal processes that will be explained in due time. My advisor, Prof. Itzhak Ben-Israel, didn't support this theme fully, but was kind enough not to care much about my work in general, as he was running for office in the Israeli parliament (and

indeed achieved his goals of becoming a member of the Knesset). My Master's thesis was a unique opportunity for me to read Kant, Descartes, Hume, Frege, Wittgenstein, as well as Einstein and Elitzur – and their viewpoints on consciousness. What became clear to me during these 2 years was that philosophy can only scratch the epidermis of such a question. If you want to actually dive into thinking about consciousness and – potentially – offer a tiny step towards understanding it, one ought to study the brain. And so I decided to do so.

ICNC

Towards the end of my 2nd year in the Philosophy of Science department in the Tel-Aviv university I joined hands with a person who shaped my academic career forever. Prof. Avshalom Elitzur from the Bar-Ilan University – who is the worst academic I have ever encountered in my short academic career, yet one of the most passionate scholars I have ever heard of (and I include in this account most of the ancient philosophers I only read about in books) – together with a small group of young academics that he had gathered from all over Israel and the world would meet weekly and secretly at my high-tech company's main office at night, after all the employees were gone, to discuss consciousness and ways to tackle it. We had a few philosophers there, 2 mathematicians, 1 biologist (who also was the only female in the group), 1 physicist, Avshalom (who is nothing but everything), and myself - who served as a computer scientists – although I never took a single class in CS, my army background in cryptanalysis proved to be immensely useful when codes and math were encountered and allowed me to serve as equal member in our “olympia academy”. We used to sit for hours a week, usually on Wednesdays, with some snacks and good spirit, film the sessions, have one or two of the members present in an informal journal-club fashion a paper that they read recently

that had to do with consciousness, and just talk. Occasionally Avshalom would bring guests to join us for these meetings – I never figured out how he afforded their visits – some of which were unknown to any of us, but required us to shift to broken English, and some of which were amazingly famous and known among the “consciousness” society that I was humbled to be hosting in our tiny office. They came because of his ability to attract people and stayed with us. Later on – having visited the Tucson consciousness conference, and the ASSC conference I encountered those people again, and to this date am unclear how Avshalom was able to make people from all over the world come attend our gatherings. I do know that I am forever grateful to Avshalom, who did the noblest thing an advisor can do to his protégée when he felt I needed it – kick me in the butt away from him. Although I was offered a very completing and attractive option to be a Ph.D. student under his mentorship at the Bar-Ilan University, in the (at the time not fully established) brain institute to be founded by Moshe Abeles (who at the time was the Israeli Guru of neuroscience who just left the Hebrew University with lots of money to found this institute at Bar-Ilan); and while I was offered a full, 4 year presidential scholarship (in Israel, Ph.D. students commonly pay full tuition for their Ph.D. work, and finish their work within 5-8 years, while I was offered the chance to do so in 4 years, and get PAID to do so!); and while it seemed that my life would be very easy having chosen Bar-Ilan and Avshalom – he was noble and kind enough to tell me, as his final advice as my mentor, that I should pack my bags and apply to the Hebrew University Interdisciplinary (a recurring motive in my life) Center for Neural Computation (ICNC). This, he said, was the “only place that will make a neuroscientist out of you and allow you to actually study what you’re interested in. You need to go through a neuroscience boot camp in order to start research in the right way, and I can’t provide that for you”. After having everything easily in my hands, or in the

words of Tyler Durden being “close to being complete” I started over. Applied to a new university and went through an acceptance committee once again.

[Naf]Tali Tishby and Eilon Va’adia were not highly impressed with me at the interview. They asked me a direct question: “Are you a marathon runner, or a 100 meter runner?”, because we don’t need “geniuses” here – we need people to spend a lifetime digging into boring questions in order to advance science step by step”. I couldn’t lie. I was not a marathon runner. I cared about the answers, not about the process. I told them that if I had known the answers to my 3 immortal questions – I would not have set foot in their office for an interview. That was all I cared about. I really don’t care about the folding of the C60, or the ability of *Drosophila* to mutate. I couldn’t lie. But somehow – it worked. Despite my not being a marathon runner at heart, Tali Tishby liked my passion and decided to give me a shot. I was accepted at the ICNC. “We can make him into a marathon runner.” Boy did they know what they were talking about....

Marathon

Nothing in my prior 4 years of physics, biology, medicine, law, film, literature, psychology, engineering, political science, and philosophy compared to my 2 years at the ICNC. In the words of David Baltimore at the commencement ceremony for Caltech: “Most of you are used to being the first in your classes. Well, I have bad news for you guys. From now on half of you are going to be below average.” And boy I was. Accompanied by 11 other brilliant students, chosen carefully from all over the country (as at that year the ICNC was at its peak and had hundreds if not thousands of applicants), I had to adhere to higher standards. Neuroscience became a very sexy topic in Israel in 2004. Institutes were opening, and my dad, who at first

didn't see the point of me attending that "unknown shelter at the Hebrew University", and going for a Ph.D. in a field that has no clear profession at the end, started seeing hope in this choice of mine. "Well - he could always resort to computer science", he told my mom, seeing the word "computation" at the end of the ICNC title. The classes were extremely intense and hard. Having taken my studies while working a full-time job in a hacking company in Tel-Aviv, having to drive 3 mornings a week to take classes, and in between fly across the globe to perform penetration tests on banks whose security systems were flawed made my life unbearable. Luckily, I had 3 things that made a difference. First and foremost — for the first time in my life I felt I had found something I was totally immersed in. I no longer cared about not being an archeologist (although upon writing these lines I do feel a tiny shiver thinking that I might end up not digging and running away from a huge rolling stone with a whip in my hand). I loved learning about the brain. Second — my classmate Jonathan Rubin made it possible for me to partake in all my activities while being a Ph.D. student by sitting long hours with me after class and explaining to me in simple words the complex ideas behind what I had just missed by having spent time breaking into Bank of America instead of listening to Chaim Sompolinsky's Neural Network class. The 3rd important anchor in making me a neuroscientist was a combination of two people: my beloved girlfriend at the time — Galit, who chose to break up with me just in time when I had to choose an advisor. I was miserable at the time, having lost my girlfriend and "love of my life" at the time, and I could not cope with the loss. I had to get as far as possible from Israel was all I knew.

That's when I started focusing on finding an advisor. I knew what characteristics my advisor should have, and I knew what types of research I was fascinated by. But finding an advisor that matches those did not seem trivial at all.

There was, however, one lab... but... that probably would be impossible... it's not even in Israel... no... it's definitely impossible... well... if I could do whatever I wanted, I would definitely go work at Christof Koch's lab at Caltech. I remembered having read his book, and the recent papers, and I thought it was something I could find fascinating, I told this to Idan Segev, who was on the ICNC council at the time. "Well... you know what? It's your lucky day. First, Christof is in fact coming here in 2 weeks as my guest to give a talk about consciousness. And second, the ICNC and Caltech have this "sister programs" fellowship. You could take your classes and qualifying exam from here and use those to apply to Caltech. With my recommendation and a clear research project under Christof - I am sure it will be possible. I wasn't as sure as Idan. But a final push from the girlfriend break-up, along with the compelling visit of Christof to the ICNC made it clear to me - I am willing to risk everything to give it a shot. But it was a very risky one. Why didn't I go to Christof's lab for a few months, take a leave of absence from my full-time job, and see if I even like it there. If he even likes me. I became greedy. From being sure that if Christof took me I would surely go work in his lab, I started interviewing him for the job of my advisor in my mind. "We'll see how he is...." The terms were easily negotiated upon Christof's visit: I will come January 1st, 2005, for a few months rotation, where I would take a project. Afterwards we would see how things evolved.

caL01jan

On January 1st, 2005, I entered Shannon's office. Christof came outside to shake my hand and also asked me who I was and what I wanted. I reminded him. Shannon and Phillippe were assigned to get me started. A computer and a username were the first on the list. Having gotten a place in the old Beckman basement lab, I was asked by Philippe the IT guy to choose a

password for my Klab account. “Something with letters, numbers, and different cases.” I chose “caL01jan”. It stood for California, January 1st. This was when I formally started my 4 years time at the best place in the world I have ever spent time in. The fastest-passing 4 years, and the most exciting marathon I have ever run.

Kelrof

Over the course of 4 years at Caltech, Christof made me figuratively and literally run marathons numerous times. From the wrecked “kid” who had started school 10 years ago, I learned that folding proteins and studying the movements of C. elegans is not necessarily a bad way to know God’s thoughts, and mostly I learned that within me lies the capacity to fully dedicate myself to one question, one notion, one riddle – giving up many others, and still being fascinated by it to the fullest. Day after day. 4 years in a row.

Dreams from my Father

My father was the one who pushed me first into academia. I can’t remember who it was that pushed me to be curious when I was a little boy, who made me want to read at a younger age, be fascinated by Geometry, spend hours with an English dictionary trying to make King Graham “Open Door” or “Kill Monster” in King’s Quests I to III which were my introduction to English, or to learn how to break assembly codes on my XT computer in the later 80s. I do remember how the notion of a father was the most important thing for me when choosing to go back to Israel after 5 months at Caltech, pack all my belongings, “fully” break up with my girlfriend, quit my job at a company which I was one of the founders of, and leave for the unknown.

It was a simple, seemingly meaningless sentence that Christof used the day before I went back to Israel to decide about my faith, as he offered me the chance to come back to his lab, this time formally as his and Caltech's student, for fulltime research. He said that, aside from all the hard work we were going to do together, aside from his being impressed with my speedy learning curve and that we already had a good lead for a research study based on my 4 months work at his lab – aside from all those, he saw that we actually got along very well. And this, to him, was a very important factor. Because, as he said, “you know, in Germany, where I did my Ph.D., they don't call the advisor ‘advisor’ or ‘mentor’. They actually use the term ‘Doctor-father’, and that is how I see the relationship between me and my students.” This was when I knew that if I chose to leave my family and friends behind to go take a leap to the unknown, where I would be away from the safe haven I carefully established in Israel over 27 years – away from Jonathan Rubin, who can help me remember the role of the Basal Ganglia 5 minutes before I am to talk about it in a Journal club, or remind me that there are only 4 “C” vertebra in the spinal cord; away from my artistic friends who did not care about my work or research even a tiny bit, and who thus forced me to learn how to explain things in layman's words and make a very marathon-like study into a 100m fascinating explanation; away from everything that was familiar and safe – if I chose to take this step, I had better do it knowing that I had some sort of a replacement family, and a “doctor-father” sounded just right.

Kelsey

Since my arrival at the U.S., 4 years ago I had found a family. Thanks to my research, and thanks to the world that was uncovered for me in sunny Los Angeles, I learned a lot about consciousness. A lot about the brain. A lot about the world. About science. About people. But

mostly — I learned about myself. These pages are not your typical Ph.D. thesis. Embedded in the coming lines are not just numerical results, but a competing story about the mind and its works. In the following chapters I will outline my story: The story of science the way I learned to see it. The story of a scientist the way I learned to see it. But mostly, these lines are a love story. A story of a person falling in love with research, with an advisor, with a country, a city, with riddles and questions, and with some answers and — like every good Los Angeles story — with a girl.



Introduction: Setting the Stage

*Do you suppose I could buy back my
introduction to you?"*

Groucho Marx

In the course of my research I used various paradigms and experimental methods as well as various techniques and algorithms for my work. The studies described hereafter were conducted over the course of four years and included very many subjects from very many types of populations. In chapter Four I will detail the question that I pursued in this thesis, and the theory that guided my thought. In chapter Five I will describe in short the background and the general relevant details pertaining to the subjects I had the fortune to work with. These include patients with epilepsy undergoing brain surgery, subjects with autism syndrome disorder, subjects with Agenesis of the Corpus Callosum, subjects with face prosopagnosia, a unique subject with no amygdala, and very many controls from Caltech, the greater Los Angeles area, and some of my friends and family. In chapter Six I will discuss the background and review the prior work done on attention and its general ideas of modeling through competition. In chapter Seven I will elaborate on the relevant prior studies of face processing and detection in general and particularly on the brain mechanisms which are involved in those as an introduction to the main stimuli I used in many of my studies, as faces are a very special type of attention-attracting entity that I studied extensively in the course of my research. In chapter Eight I will introduce the three major methods I used for my work: single neuron recordings from the brains of the epilepsy patients; eye-tracking as well as other psychophysical methods with the remaining populations; and computational modeling of brain interactions using either neuron network representation of the mechanisms behind the working of my epilepsy patients' attention allocation methods, or a saliency model which is used to model the eye-tracking attention allocation. In chapter Nine I will describe a psychophysical study done with subjects at Caltech showing that people looking at images and asked to rate how interesting they are turn out to be very consistent in their judgment metrics —

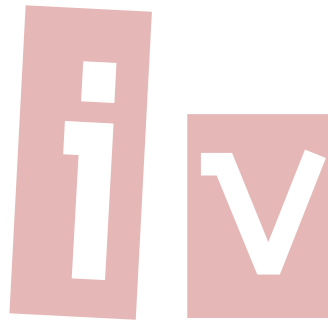
both with themselves and also with other independent viewers. This turns out to be true even if the images contain very little meaningful content, are presented very briefly, or if the same subject's answers are compared a year apart. In chapter Ten I will describe an experiment showing that, when presented with faces, subjects tend to be very similar in their attentional allocation – they basically look at the faces quickly, and spend most of their time doing so. In chapter Eleven I will present a study where we extended the prior work on faces, to another entity which we found relevant and important – text. Again, we show that high-level entity wins the competition for subjects' attention regardless of task and. This leads to the final study in that series where we - in chapter Twelve - show that when subjects try to avoid looking at either a face or text, they fail to do so in many ways because those are so good in winning our attention. In this study we suggest that human attraction to faces – face attractiveness – is an innate mechanism rather than an acquired one. In chapter Thirteen I will discuss a different novel approach we had to looking at the innate mechanisms in our brain that lead to our choosing what to look at. A closer look at the competition as it is reflected by the early saccades our eyes make to objects in the scene, and the interplay between bottom-up and top-down mechanisms in guiding the gaze. Based on these studies, I present in chapter Fourteen an extensive model of attention with an additional face component that better predicts the fixations of subjects based on the prior studies. This model predicts the data from our previous studies with a high degree of accuracy. This is especially true during the early stages of viewing, suggesting that an algorithm based on features and content of an image could predict the brain's allocation of attention to a natural scene. In chapter Fifteen I will expand the model to now include multiple high-level entities, and show that altogether these typically show an increased accuracy in predicting observers' gaze. Additionally, in chapter Sixteen I will show a

novel way to test the ability of a model to accurately predict what people see. A model which takes the scanpaths of various subjects and uses that to try to identify the image they were looking at. In chapter Seventeen I will detail a study performed with subjects with autism. Given the high attractiveness of social cues among our subjects in the previous studies, we wanted to test the hypothesis that individuals with less interest in social scenes – such as people with autism – would show different behavior under similar conditions. The study described in chapter Seventeen even suggests that there are differences in viewing patterns between autistic individuals and controls. In chapter Eighteen I will briefly mention various populations of subjects I tested using the same paradigms in order to identify population differences. These are subjects with agenesis of the corpus callosum, amygdala-lesion subjects, subjects with face blindness (prosopagnosia), and various anecdotal cases of outliers with interesting results, such as two identical twin sisters with almost identical ratings in a task they performed separately, and a similar pairing between an autistic child and his mother. Chapter Nineteen will take the study to the realms of its applications and the actual uses of such models of attention. I will describe a suggested usage of our works to the fields of business and marketing, and the relevance of neuroscience to fields which to date are not benefiting from it enough. In chapter Twenty I will first discuss the study of attention and competition between brain regions using single neuron recordings with epilepsy patients. In this study we projected the thoughts of subjects on a computer screen in real-time. As a paradigm we had subjects try to move an image on the screen with their thoughts alone. As a competition metric we had them move away from one image and closer to another one, while both were on the screen, and were governed by competing brain regions. In chapter Twenty One I will describe a computer model based on neural network representation of the neurons involved in the task described in

chapter Nineteen. The model suggests an explanation and a quantitative measure of the involvement of attention versus imagery in the task.

In chapter Twenty Two I will describe three unique patients who were able to not only reach perfect control of their neuronal activity on demand, but were actually able to use that ability to play a computer game using mind-control. We designed a special version of the famous “space invaders” game and had these patients play it purely by altering their thoughts.

I will end in chapter Twenty Three by discuss future direction and questions that should be pursued following my work, and by concluding the work in chapter Twenty Four with ideas of how this work plays a role in the overall questions that science is baffling with nowadays.



A question and a theory

First, you know, a new theory is attacked as absurd; then it is admitted to be true, but obvious and insignificant; finally it is seen to be so important that its adversaries claim that they themselves discovered it

William James

Every thesis, I believe, should have a main question that the authors are interested in answering, and a theory or method to answer it. In the coming pages I will detail the questions that guided me in my research along the four years of studies at Caltech, and the theory and method of inquiry that I utilized in order to answer them. First let me point out the main question.

The question that I was interested in during my research is **whether we are able to control which thoughts and sensations we become conscious of**. While throughout our lives we are aware of very many things, and we suppress equally large amount of things, we normally do not control in a conscious way which of those we would become aware of. We can control our behavior, to some extent, and we can choose what to expose ourselves to, and as such what to let our brain encounter, but altogether many of our activities are the outcome of internal states that we cannot necessarily access. If we see a room full of people, we can choose to attend to one person, to focus on one side of the room, or to ignore a particular aspect of the scenes, but we cannot control what eventually enters our brains. Some sounds, smells, and feelings that arise from the environment will penetrate regardless of our choice. These sensations would reach our consciousness and affect our behavior, but we are not able to select them. Our brains act as a ‘sponge’ of information that is inputted and acted-upon without us having much control of it. Throughout my studies I was interested in understanding how our brain selects the parts of the scene that we are aware of eventually.

As a preliminary question I asked myself - as I came to study this question - if these mechanisms which I am to study are ‘reliable’ in a sense and can be studied at all. By reliable I

meant - **do the brain mechanisms which underlie our behavior work in such a way that under the very same conditions they would yield the same outcomes?**

As a method I was seeking in the beginning of my studies to find a set of experiments and stimuli that would be such that they produce repeatable results. Many experiments that rely on human behavior are prone to suffer from this caveat of our psyche - we tend to change rapidly. Luckily, at the course of my studies I stumbled upon a set of conditions and stimuli that let a subject select his behavior subjectively yet reflect his attentional choices in a replicable manner.

I am a 'classical physics person' in my way of thinking. That is, I hope and work as if the world has a set of governing rules and mechanisms which can be explained. I do not believe in god, and I am the least spiritual person I know. If I discover a sense of superstition in me, I immediately eradicate it, in order to not allow any access to non-logic in my life. This, at least, is my method of scientific inquiry. I do, however, understand cognitively and profoundly that the world is 'quantum', that is - based on statistics and on various effects that cannot necessary be repeated on a trial by trial basis. In the course of my studies I had to repeatedly let go of my inner, deep and robust method of thinking of the world as black or white. While I tend to think of law and order as the ultimate level of ethics, I had to agree that many times these require interpretation. While I tend to see negatively hypocrisy and unfairness - rules that apply to one but not the other - I had to accept the fact that the nature of the world is that it sometimes calls for differences in the approaches to a solution by two people. And while I thought that the world of science should be made of rules, mathematical structures and non-abstract formal definition to its questions and arguments, I had to accept that the science of the brain many times does not allow for those to apply. Biology is lying somewhere in the grey

zone, and I had to adapt my methods of thinking to a world that resides in the grey and let loose of my black-and-white classical method of thinking many times. A person could be diagnosed by one as autistic and by another one as not. A person can think he is performing the very same actions, but get different results when deep brain processes are involved. An explanation can work perfectly for 99% of your subjects and fail miserably with the remaining 1%, still being regarded correct by the scientific community.

As such I had to choose a method of research that will answer my question, allow me to believe the results - despite their being statistical, and be important and interesting for me to ponder about.

As I see myself as a person who is very good in interacting with people and understanding them through simple interactions I decided early on that I would prefer to do my research with humans (rather than animals, which I personally think I cannot work with for my personal ethical reasons) and with modeling. The ideal output of each study would be a computation model which predicts human behavior. Being interested in psychiatric disorders and in disorders in general - as they reflect an extremity of the human psyche - I selected to work with populations of people whose conditions include epilepsy, autism, lack of understand of emotions, etc. I was lucky to have access to these people who were happy to work with me and excited to answer my questions and let me dive into their lives.

In order to answer my questions I conducted a set of studies which step by step come to target my main question. In the first study I report here I tried to see how reliable people are in answering questions and what type of stimuli would make people be most similar to others. The ideas was to find a set of questions, experimental conditions and stimuli that would benefit

future experiments of mine by making the variability across subjects as smaller as possible. I was able to learn that asking people to judge their interest in scenes, and making the scenes natural and similar to those they would encounter in their environment makes the variability small. I later learned that having a social entity like a person in the scene decreases the variability even more. I used human faces and people's viewing patterns to learn about the things they care about the most, and the earliest. I figured that in our everyday life we can see the things we become aware of as the output of an internal competition in our brain. Many sensory inputs penetrate our eyes, but we end up noticing only few. All the pixels of the environment go through our retina, but as they become concepts in our mind, we select only few of them and become conscious of those. I therefore decided that a method of inquiry that would involve generating these types of competition in our brains, in a controlled fashion would allow me to tap into the mechanisms that select what to become conscious of and what to suppress. Eye-tracking of multiple attractive objects was one way to address this competition as putting two attractive objects and seeing which one wins our attention, to what extent and how often can show us what happens in our brains when we are exposed to those. I conducted a series of experiments along this method of inquiry and indeed was able to produce a computational model of our fixations patterns showing that people are in fact similar to each other when seeing natural scenes, and that our attentional shifts can be predicted by an algorithm.

As a second method I used recordings from the brains of humans, to access the competition directly. I created an experiment where subjects were instructed to alter their brain activity to make one or another stimulus win the competition and was therefore able to reliably understand to what extent we can control the flow of information and the competition in our

brain. Again, after learning that this competition is accessible and can be altered by our own will, I used the data from our subjects to generate a model of the mechanisms in our brain that modulate the competition for our attention.

Finally, I was able to use the data and a set of additional experiments and studies conducted with populations of subjects who ‘play the game differently’. As these people yield very different outputs under the same conditions (either, not looking at what others typically look the most, or being unable to access the mechanisms we commonly use to win competitions such as fear or cognition, or other such unique differences) I could learn what brain mechanisms are likely to govern these competitions. Lack of one brain region, for instance, that leads to a clear deficit in a task suggests that this region is involved in performing it. By working with these populations I could add to my model and experimental data some hypothesis on the way by which our brain selects what to become conscious of, why these things are selected, how rapidly and how frequently these selections occur. In this work I show that in fact as experimenters we can control to some extent the attention of our subjects purely by selecting what to expose them to, and that we can guide our subjects with clear feedback on their behavior to learn how to control what they are conscious of and what thoughts dominate their attention.



Subjects

I've learned that our background and circumstances may have influenced who we are, but we are responsible for who we become.

James Rhinehart

Let me first describe in brief the populations with which I had the fortune to work during my research and what's known of their conditions and behavior. Aside from the very many controls which came from the Caltech population of poor undergrads in dire need of money and able to sustain hours of looking at rotating dots and moving bars, being food-deprived, or shocked for a \$15 prize, I used my friends, family members, random visiting lab guests, students of classes I taught, and an ample amount of Craigslist volunteers. Outside of the control group, I used subjects with stereotypical brain disorders with which I could test various hypothesis concerning their condition. These were subjects with Autism Syndrome Disorder (ASD) (some of which had other variants of behavioral disorders such as anxiety, depression or various emotional disorder that were not studied by me and therefore won't be discussed here); individuals with Agensis of the Corpus Callosum (AgCC); individuals with epilepsy; a single subject with a rare disorder leading to amygdala lesion; and two subjects with face-blindness (Prosopagnosia).

Autism

- “In th next task I want you to try and avoid looking at that face.”
- “Which face?”
- “This face. The one in the center of the screen!”
- “Oh... I didn’t even notice it. I was looking at this cool Rubik’s Cube on the table.”

Autism as a syndrome was first described by Leo Kanner, a child psychologist, in 1943. His initial description, based on 11 case studies emphasized “...an innate inability to form the usual, biologically provided affective contact with other people”. For a long time, autism was thought to be a consequence of bad parenting, and the “refrigerator mother” theory (Bettelheim, 1972) lasted from the 1950s well beyond the 1970s. Bernard Rimland (Rimland, 1964) and Michael Rutter (Rutter, 1968) established empirically that the parents of autistic children were not different in their parenting from the parents of non-autistic controls, and helped build a case for a neurobiological basis of autism, although the relationship between parents and autism kids is still being studied. In this work I will discuss one such correlation between a mother and her autistic kid who show higher than normal correlation in their responses to a psychophysics experiment. Note that this correlation might not be the cause but rather the effect. Parents of kids with autism tend to need to care for them even more than usual parents and hence get to “know them even better”. Autism is now recognized as a neurodevelopmental disorder

manifesting within the first 3 years after birth and progressively worsening in the course of life. The core symptoms are impairments of sociability, communicative skills, and imagination, together with stereotypic behaviors and repetitive tendencies (All, 1980). At the cognitive level, all autistic children seem to display some form of abnormality in perception, attention, and memory (Ben Shalom, 2003; Dakin & Frith, 2005; Sanders, Johnson, Garavan, Gill, & Gallagher, 2008).

Genetic analyses have revealed that autism is a polygenic disorder that any one or more of a set of genes can predispose toward, but no one gene has been found to cause autism (Bonora, Lamb, Barnby, Bailey, & Monaco, 2006; Cook Jr., 1998; Lamb, Moore, Bailey, & Monaco, 2000; Persico & Bourgeron, 2006). The primary cause of autism is most likely a form of epigenetic alteration during development (Beaudet & Zoghbi, 2006) which triggers a cascade of diverse neuropathologies depending on the timing of the epigenetic attack.

Autism encompasses a spectrum of disorders ranging from severe mental retardation to high functioning Asperger's and "idiot savants", with many brain regions implicated, making it difficult to develop a unified theory of autism. High-functioning autistics, who are mainly the ones I performed my studies with, have been viewed as the exception to the mainstream view that autism is a severe form of mental retardation with poor cognitive capabilities (Pring, 2005).

Neuropsychology

Two major categories of cognitive theories have been proposed about the links between autistic brains and behavior. The first category focuses on deficits in social cognition. Hyper-systemizing hypothesizes that autistic individuals can systematize — that is, they can develop internal rules of operation to handle internal events — but are less effective at empathizing by

handling events generated by other agents (Baron-Cohen, 2006). It extends the “extreme male brain theory”, which hypothesizes that autism is an extreme case of the male brain, defined psychometrically as individuals in whom systemizing is better than empathizing (Baron-Cohen, 2002). This in turn is related to the earlier theory of mind, which hypothesizes that autistic behavior arises from an inability to ascribe mental states to oneself and others. The theory of mind is supported by autistic children's atypical responses to the Sally-Anne test for reasoning about others' motivations (Baron-Cohen, Leslie, & Frith, 1985), and is mapped well from the mirror neuron system theory of autism (Iacoboni & Dapretto, 2006). The second category focuses on nonsocial or general processing. This is the one we focused on in our research. Executive dysfunction hypothesizes that autistic behavior results in part from deficits in working memory, planning, inhibition, and other forms of executive function (Kenworthy, Yerys, Anthony, & Wallace, 2008). Tests of core executive processes such as eye movement tasks indicate improvement from late childhood to adolescence, but performance never reaches typical adult levels (O'Hearn, Asato, Ordaz, & Luna, 2008). A strength of the theory is predicting stereotyped behavior and narrow interests (Hill, 2004); two weaknesses are that executive function is hard to measure (Kenworthy et al., 2008) and that executive function deficits have not been found in young autistic children (Sigman, Spence, & Wang, 2006). Other theories, such as the weak central coherence theory, hypothesizes that a limited ability to see the big picture underlies the central disturbance in autism. This theory is predicting special talents and peaks in performance in autistic people (Happé & Frith, 2006). Another related theory – enhanced perceptual functioning – focuses more on the superiority of locally oriented and perceptual operations in autistic individuals (Mottron, Dawson, Soulieres, Hubert, & Burack, 2006).

Diagnosis

Diagnosis is based on behavior, not cause or mechanism (Baird, Cass, & Slonims, 2003). Autism is defined in the DSM-IV-TR as exhibiting at least six symptoms in total, including at least two symptoms of qualitative impairment in social interaction, at least one symptom of qualitative impairment in communication, and at least one symptom of restricted and repetitive behavior. Sample symptoms include lack of social or emotional reciprocity, stereotyped and repetitive use of language or idiosyncratic language, and persistent preoccupation with parts of objects. Onset must be prior to age three years, with delays or abnormal functioning in either social interaction, language as used in social communication, or symbolic or imaginative play.

Several diagnostic instruments are available. Two are commonly used in autism research: the Autism Diagnostic Interview-Revised (ADI-R) is a semistructured parent interview, and the Autism Diagnostic Observation Schedule (ADOS) uses observation and interaction with the child. The Childhood Autism Rating Scale (CARS) is used widely in clinical environments to assess severity of autism based on observation of children (Volkmar, Chawarska, & Klin, 2004).

At Caltech we have administered to the recruited autism candidates the following diagnostics:

ADOS and ADI-R.

ADOS

The Autism Diagnostic Observation Schedule is a standardized protocol created in 1989 for assessing social and communicative behavior associated with autism. The protocol consists of a series of structured and semi-structured tasks that involve social interaction between the examiner and the subject. The examiner observes the subject's behavior and assigns identified

segments to predetermined observational categories. Categorized observations are subsequently combined to produce quantitative scores for analysis. Research-determined cut-offs identify the potential diagnosis of autism or related autism spectrum disorders, allowing a standardized assessment of autistic symptoms.

ADI-R

The Autism Diagnostic Interview – Revised is structured interview conducted with the parents of individuals who have been referred for the evaluation of possible autism or autism spectrum disorders. The interview can be used for diagnosis purposes for anyone with a mental age of at least 18 months and measures behavior in the areas of reciprocal social interaction, communication and language, and patterns of behavior.

The interview covers the referred individual's full developmental history, is usually conducted in an office, home, or other quiet setting by a psychiatrist or other trained and licensed professional, and generally takes one to two hours. The caregivers are asked 93 questions, spanning the three main behavioral areas, about either the individual's current behavior or behavior at a certain point in time (Lord, Rutter, & Couteur, 1994). Because the ADI – R is an investigator-based interview, the questions are very open-ended and the investigator is able to obtain all of the information required to determine a valid rating for each behavior.

The first section of the interview assesses the quality of social interaction and includes questions about emotional sharing, offering and seeking comfort, social smiling, and responding to other children. The communication and language behavioral section investigates stereotyped utterances, pronoun reversal, and social usage of language. The restricted and repetitive behaviors section includes questions about unusual preoccupations, hand and finger

mannerisms, and unusual sensory interests. Finally, the assessment contains questions about behaviors such as self-injury, aggression, and over-activity which would help in developing treatment plans.

The interviewer determines a rating score for each question based on their judgment of the caregiver's response.

0 = "Behavior of the type specified in the coding is not present".

1 = "Behavior of the type specified is present in an abnormal form, but not sufficiently severe or frequent to meet the criteria for a 2".

2 = "Definite abnormal behavior"

3 = "Extreme severity of the specified behavior"

7 = "Definite abnormality in the general area of the coding, but not of the type specified"

8 = "Not applicable"

9 = "Not known or asked"

A total score is then calculated for each of the interview's content areas. When computing the algorithm, a score of 3 drops to 2 and a score of 7, 8, or 9 drops to 0. An autism diagnosis is indicated when scores in all three behavioral areas exceed the specified minimum cutoff scores.

Our subjects used the following cutoff scores:

Social interaction = 10

Communication and language = 8 (if verbal) or 7 (if nonverbal)

Restricted and repetitive behaviors = 3.

On top of the two tests mentioned above, some subjects were tested with various other diagnoses such as the Benton, STAI, and the WAIS (IQ).

Benton

The Benton Visual Retention Test is an individually administered test for ages 8 – adult that measures visual perception and visual memory. It can also be used to help identify possible learning disabilities. The child is shown 10 designs, one at a time, and asked to reproduce each one as exactly as possible on plain paper from memory. The test is untimed, and the results are professionally scored by form, shape, pattern, and arrangement on the paper.

STAI

The State-Trait Anxiety Inventory is an instrument for measuring anxiety in adults. The STAI differentiates between the temporary condition of "state anxiety" and the more general and long-standing quality of "trait anxiety". The essential qualities evaluated by the STAIS scale are feelings of apprehension, tension, nervousness, and worry. Scores on the STAIS scale increase in response to physical danger and psychological stress, and decrease as a result of relaxation training. On the STAI, consistent with the trait anxiety construct, psychoneurotic and depressed patients generally have high scores.

WAIS

Wechsler Adult Intelligence Scale is a general test of adult intelligence (i.e., an IQ test), first published in February 1955 by David Wechsler; the fourth and most recent edition of the test

(WAIS-IV) was released in 2008 by Pearson. Wechsler defined intelligence as “The global capacity of a person to act purposefully, to think rationally, and to deal effectively with his/her environment” (Wechsler, 2007).

There are four index scores representing major components of intelligence:

- Verbal Comprehension Index (VCI)
- Perceptual Reasoning Index (PRI)
- Working Memory Index (WMI)
- Processing Speed Index (PSI).

Two broad scores are also generated, which can be used to summarize general intellectual abilities:

- Full Scale IQ (FSIQ), based on the total combined performance of the VCI, PRI, WMI, and PSI
- General Ability Index (GAI), based only on the six subtests that comprise the VCI and PRI.

Other than the additional diagnosis tests abovementioned, subjects were treated exactly the same as our control subjects and performed the same tasks under the same conditions.

Relevance to attention study

People with autism do not attend to the same things we do. If they do end up attending to some things like us, they do it less frequently, more slowly, and most likely for different

reasons. I used the autism group in this study as a variant of the attention study, wanting to see if things we attend to almost involuntarily because of our upbringing or nature will have similar effects on individuals with autism who don't "care" about the same things necessarily. Particularly, I focused in this study on social cues such as faces, trying to see if they would result in different attentional allocation measures.

AgCC

Agenesis of the Corpus Callosum is a complete or partial absence of the corpus callosum throughout development. As the corpus callosum is the widest bridge of information between the two hemispheres in our brain, this entails significant problems in flow of information between the two sides of the brain. While famous studies by Gazzaniga (Gazzaniga, 1970) with patients whose corpus callosum was cut as a severe method of treatment for recurring epilepsies, showed an inability to transfer information resulting in mistakes of perception, the AgCC tend to have better transfer of information since the brain is able to compensate for the loss of connectivity due to plasticity in early childhood.

Competition

The study of subjects with lack of information transfer between the two hemispheres surely affects the speed and ability to analyze information. The split brain patients are the ultimate example of clear competition on the attention and resources within a single person. They manifest that evident competition strongly when given a task that requires both sides of the brain in a competing method. A classical example of the competition among these patients is shown with a split-brain patient who is given a task to solve a Tangram puzzle (where visual image is to be formed using 7 pieces in different shapes). If the patient is using only his left hand for the task and views the puzzle with his right eye alone, he performs the task perfectly. When told to not use his left hand, but try to perform the task with his right hand alone – not getting accurate information from the corresponding left hemisphere – he fails miserably. Interestingly, the competition is nicely manifested in this example when repeatedly the left hand (on which he sits, in order to avoid using it) occasionally “escapes” his guard and comes

to the rescue of the right hand; until he is reminded by the experimenter again to not use it. When the patient is asked to use both hands for the task the competition becomes even more evident when the left hand easily starts performing the task, while occasionally the right hand disturbs it and hurts the performance by moving correct pieces from their location or otherwise introducing mistakes into the task. (See video at <http://www.youtube.com/watch?v=0lmfxQ-HK7Y>.)

This makes AgCC patients a unique case of separation between the hemispheres, very interesting candidates for the study of attention as competition between brain regions.

Background

AgCC is a complex condition, which can result from disruption in any one of the multiple steps of callosal development, such as cellular proliferation and migration, axonal growth, or glial patterning at the midline. Current evidence indicates that a combination of genetic mechanisms, including single-gene Mendelian mutations, or single-gene sporadic mutations might have a role in the aetiology of AgCC. Studies report that 30-45% of cases of AgCC have identifiable causes. Approximately 10% have chromosomal anomalies and 20-35% have recognizable genetic syndromes (Bedeschi et al., 2006). However, if we only consider individuals with complete AgCC, then 75% of cases of complete AgCC do not have an identified cause (Paul et al., 2007).

One hospital-based study reported that just under a third of patients with AgCC were developmentally “normal” or only mildly delayed (Shevell, 2002). A longitudinal study of 17 children prenatally diagnosed with AgCC showed that nearly all patients had at least mild behavioral problems (Moutard et al., 2003). This suggests that isolated AgCC, even when not

ascertained clinically, still causes behavioral and cognitive impairment. Parents often report that when their child was diagnosed with AgCC, they were told that the prognosis was unclear, ranging from severely delayed to “perfectly normal”. However, as more individuals with primary AgCC are identified and assessed with sensitive standardized neuropsychological measures, a pattern of deficits in higher-order cognition and social skills has become apparent even in the so-called “normal” individuals with AgCC.

Behavior

AgCC has a surprisingly limited impact on general cognitive ability. Although the full-scale IQ can be lower than expected based on family history, scores frequently remain within the average range (Chiarello, 1980). In an unexpectedly large number of persons with primary AgCC (as many as 60%), performance IQ and verbal IQ are significantly different (Chiarello, 1980; Lasonde & Jeeves, 1994). Impairments in abstract reasoning (David, Wacharasindhu, & Lishman, 1993), problem solving (Fischer, Ryan, & Dobyms, 1992; Imamura, Yamadori, Shiga, Sahara, & Abiko, 1994), generalization (Solursh, Psych, Margulies, Ashem, & Stasiak, 1965) and category fluency (the ability to list multiple items that belong to a semantic category, for example, names of animals) (David et al., 1993) have all been consistently observed in patients with primary AgCC. While neuropsychological research into domains such as memory, attention and spatial skills is under way in large samples of patients with primary AgCC, currently published results in these domains are limited to a few case studies that do not yet provide consistent findings. Parents of individuals with AgCC consistently describe impaired social skills and poor personal insight as the features that interfere most prominently with the daily lives of their children (Badaruddin et al., 2007; Lasonde & Jeeves, 1994; Stickles,

Schilmoeller, & Schilmoeller, 2002). Specific traits include emotional immaturity, lack of introspection, impaired social competence, general deficits in social judgment and planning, and poor communication of emotions (for example, individuals prefer much younger friends, have a marked difficulty generating and sustaining conversation, take all conversation literally, do not take perspective of others, and are unable to effectively plan and execute daily activities such as homework, showering, or paying bills (Stickles et al., 2002)). Consequently, patients with primary AgCC often have superficial relationships, suffer social isolation, and have interpersonal conflict both at home and at work due to misinterpretation of social cues. Responses of adults with primary AgCC on self-report measures also suggest diminished self-awareness. The patients' self-reports are often in direct conflict with observations from friends and family. One potential factor contributing to poor self-awareness may be a more general impairment in comprehension and description of social situations. For instance, when presented with highly provocative social pictures (for example, photos of mutilations), adults with AgCC tend to underestimate the emotional valence and intensity of the pictures, particularly for negatively valenced pictures (Paul, Schieffer, & Brown, 2004). Taken together, the neuropsychological findings in primary AgCC highlight a pattern of deficits in problem solving, in social pragmatics of language and communication, and in processing emotion.

Are these “split-brain” patients?

Despite the lack of transfer of early visual information, individuals with AgCC display a normal ability to make comparisons of simple and easily encoded stimuli, indicating an intact interhemispheric transfer of simple or familiar information. Information can be transferred between the hemispheres in AgCC. Historically, most research with patients with AgCC

focused on the consequences of callosal absence, with the expectation that patients with AgCC would exhibit a “disconnection syndrome” similar to that seen in commissurotomy patients (Sperry, Schmitt, & Worden, 1974). The classic disconnection syndrome involves the complete lack of interhemispheric transfer and interhemispheric integration of sensory and motor information presented independently to each of the hemispheres (Sperry et al., 1974), with surprisingly subtle behavioral consequences in everyday life. As shown by visually evoked potentials, there is a complete lack of interhemispheric transfer at the level of early visual processing in AgCC (Brown, Jeeves, Dietrich, & Burnison, 1999). The hemispheric disconnection of the primary visual system in patients with AgCC results in a unique pattern of deficits in laboratory tasks that involve comparisons across the two visual fields: intact comparisons of simple stimuli and impaired comparisons of complex stimuli. One theory to explain the preserved capacity for interhemispheric transfer of simple stimuli in patients with AgCC is that simple information can be transferred via other connecting pathways, such as the anterior commissure. Structural and functional exploration of these alternate pathways for interhemispheric transfer is a crucial frontier in AgCC research.

Relevance to attention

The deficits in social communication and social interaction in patients with primary AgCC overlap with the diagnostic criteria for autism (from the Diagnostic and Statistical Manual of Mental Disorders, fourth edition; DSM-IV). Furthermore, people with primary AgCC may display a variety of other social, attentional, and behavioral symptoms that can resemble those of certain psychiatric disorders. Therefore, the AgCC group was used as a test bed for variants in the attention allocated to social cues, similar to the way in which the autism group was tested.

More so, as mentioned above, the AgCC subjects are a variant of split-brain patients with an evident measure of competition between the two hemispheres, which is an ideal test for race between resources on the single percept or behavior.

Amygdala lesion

- “But what does it mean, you can’t do a scared face?”
- “I am telling you. I just don’t know what fear is. I am never scared.”

SM is a very unique woman. She is in her early 40s and at first might look like a very typical mid-western woman, albeit a little older than her age. A closer look, though, will reveal a somewhat childish look and a very curious and engaging gaze. She is less afraid of the first encounter than a typical person. She acts after a very short interaction like a person who knows you for a while would. She hugs and holds hands with people fairly quickly, and the somewhat structured boundaries between males and females, and between familiar and unfamiliar people, are rapidly broken with her. Luckily, the same is true for me. The week I spent with SM was by far the most interesting and engaging week I have had throughout my entire 4 years of research at Caltech.

SM suffers from Urbach-Wiethe disease, a condition that caused a nearly complete bilateral destruction of the amygdala, while sparing hippocampus and all neocortical structures, as revealed by detailed neuroanatomical analyses of her Computed Tomography and Magnetic Resonance Imaging scans (see Figure 1).

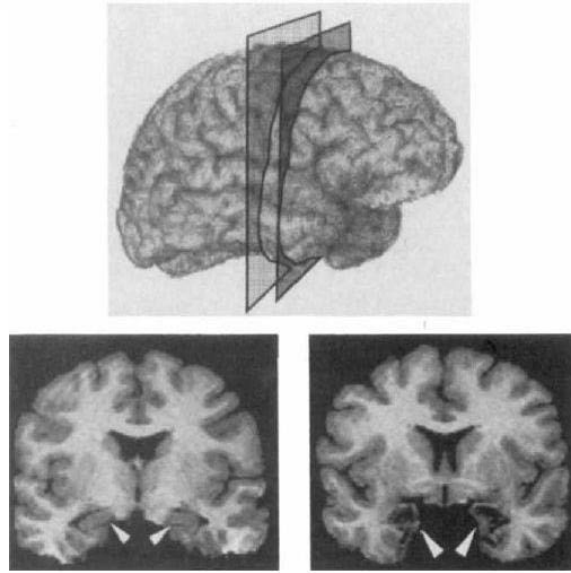


FIG. 1 t_1 -weighted MR images of S.M.'s brain. Planes of section are shown at the top, on a three-dimensional reconstruction²⁶ of S.M.'s brain. There is extensive bilateral amygdala damage (lower right image, large arrowheads) with sparing of neocortex and hippocampus (lower left image, small arrowheads). The tissue of the amygdala has been replaced by mineral deposits as a result of Urbach-Wiethe disease¹¹.

Figure 1 – Subject SM's CT scan

From (R. Adolphs, 1994)

Prior studies with SM showed a decreased performance in her ability to identify fearful emotions in pictures shown to her (R. Adolphs, 1994), and the lack of attention allocation to the eyes when trying to identify these emotions (R Adolphs et al., 2005). More so, various trials where SM was to encounter fearful emotions resulted in inability to respond to these in a normative way. I conducted a long interview with SM pertaining to that aspect of her understanding of fear, which shed light on her compensation over the lack of fear understanding, with a set of cognitive mechanisms made to suggest alternatives for the reason she does not perceive fear normatively (see Appendix 3 for a transcript of the conversation).

Relevance to attention

Clearly, SM looks at things differently. Out of a vast number of objects and resources in the environment, she chooses to look at things based on her interior set of saliency metrics that do not go hand in hand with these of controls. While a major part of our attention allocation mechanisms are made to rapidly select the most important stimulus to tell us if we are in danger (a rapidly moving object heading in our direction would win a race to our single percept rapidly, as it will be identified as potential danger). SM does not share this need. As fear is a guiding mechanism in attention selection, SM is a good candidate for the study of attention, as fear is not one of her major attractors for attention. Studying the objects which SM looks at rapidly, and the reasoning behind her attention allocation mechanisms is one of the most controlled ways to study attention, as the absence of the amygdala, which is regarded as a saliency mechanism for emotions, sheds clear light on the importance of the amygdala in driving our attention mechanisms, and the mechanisms that govern our attention mechanisms. We use SM's lack of amygdala to define our attention metrics that later are tested in the amygdala competition theories developed with our epilepsy patients. More so, as I will demonstrate in chapter 18, the amygdala plays a key role in allocating attention to faces or other variants of social scenes, and by that, the lack of amygdala will be used as a control for the level of attention allocated to scenes due to the emotional saliency in the brain, rather than pure control using bottom-up mechanisms.

Prosopagnosia

Blink

Malcolm Gladwell is a skinny 45-year-old male. He is 5'7 tall, with a very distinguished big curly hair. His skin is fairly dark, being half African-American, and his eyes are very round and deep. In fact, his eyes look like the eyes of a kid who just discovered a new thing. Very open and very excited. When he speaks, his head is tilted a little to the ground, and even though he might be shorter than you he speaks with his face looking a little down as if you're shorter than him. Malcolm Gladwell's face is very recognizable and not easy to mistake once you've seen him once for more than a brief moment, and his features are very salient and noticeable.

However, the patient Mike kept mistaking me for Malcolm Gladwell. In his mind the two of us were identical. He could not find any distinguishing visual features between us. Until I started speaking. When I spoke it was easy for Mike to tell that my non-American accent doesn't sound at all like Gladwell's and he could tell us apart.

When Mike sees a picture of Malcolm Gladwell on a computer screen, and carefully looks at him, and then diverts his eyes to me – he cannot tell the difference between us two. I showed Mike a set of dozens of pictures of Malcolm Gladwell while I was standing next to him and asked him to see if there is any way for him to distinguish us two. He could do that by picking very simple features and focusing on them for a while, but after a short time when he – again – was asked to tell which picture out of a set of images was mine and which was Gladwell's he kept making sequences of mistakes. Out of a set of 20 pictures of Malcolm Gladwell, he

identified 9 of them as me and 11 as him. Almost exactly the chance level. See Figure 2 for a picture of Malcolm Gladwell and yours truly.

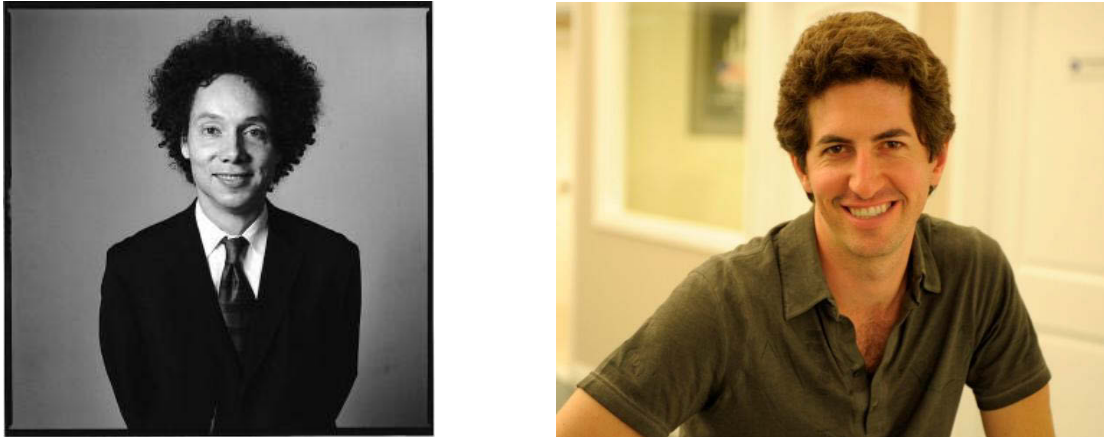


Figure 2 - Pictures of two people who seem identical to a subject with Prosopagnosia

Left. Malcolm Gladwell. **Right.** Moran Cerf. Pictures were chosen such that they make the highest resemblance among the two, according to an unbiased observer who said that when my hair is not combed there's "some resemblance in the messiness" (Kelsey Laird, personal communication). Yet, it is evident that most readers are able to tell the two identities apart.

Cats

Katy is a 25-years-old graduate student. She did her undergraduate studies at the University of California at San Diego, majoring in Chemistry. Throughout her studies Katy repeatedly found it hard to get help from her peers. Mainly, she could not ask any of them to borrow their notebooks. The reason was she just couldn't figure out who to return the notebooks to. She kept not recognizing the people she earlier spoke with and borrowed the assignments from. When she first told her mother about her problem recognizing her peers, her mom didn't believe her. The initial reaction was to suggest Katy is a racist. The chemistry classes at UCSD

include very many Asian girls, about Katy's age. Katy's mother accused Katy of not paying enough attention to learning who her classmates are, and thinking that "all Asians look alike". Katy made an effort to try and remember distinct features in her classmates, focusing on hairstyle and clothing style, but the girls in her class kept changing their styles, making it impossible. "I never borrowed people's notes unless it was by email, because otherwise I didn't know who to return the notebook to. Email is easy – you just hit 'reply'".

Katy and Mike each cope with this unique condition differently. Katy chose to be "nice to everyone" until proven otherwise. If you approach Katy on the street and start talking to her, she would answer fondly and act as if she knows you well. She doesn't recognize you at all, but will assume that if you spoke with her you must know her from somewhere and it's up to her now to gather enough evidence to figure you out – literally. With me it was easy, she said. "You have a very evident accent, although it's not a classical language". My accent is a combination of French and Israeli – which makes it harder for Katy to place. "But that's not the only thing I will use", she adds. "I know by now exactly what's unique about you". "Is it my celebrity look-alike Malcolm Gladwell", I ask based on my encounter with Mike. "Not at all", she answers, suggesting that each Prosopagnosia subject has a unique set of cues. "It's your posture. I would never mistake your posture for anyone else's." I have been with myself for as long as I remember, but I never felt or imagined there was something unique about my posture. I still don't know what it is. But for Katy, it's just enough. In an experiment where she was looking at thousands of pictures on a computer screen, very many of them of me. She could not, at first, identify me from the others. No matter how many repetitions were there she could not perform this task that seems trivial to us. She did, however, find many hints that others would not look at. "All the images you have were taken in the same day," she pointed

out. “You are the guy wearing the white shirt with his hair tied down.” Now the task was easy for Katy. But in block 3 the images were actually taken on a different day. What would happen now? At the end of block 3 Katy approached me with a blush. “Were you in block 3 at all?” she asked. “Yes, I was.” “Did you have a very short hair in that block, or was it someone else?” “It is not me. It’s my friend Yadin”. “Oh...”, she smiles and sighs a bit relieved, “good... I found this guy very attractive, and didn’t want to embarrass you by suggesting I am attracted to you.” “It’s ok,” I answered. “You’re not the first one. My friend Yadin is actually a model... you’re supposed to be attracted to him!” She feels better now. But this raises an even more interesting question – Katy can be visually attracted to someone, know that she indeed is attracted to him, but have no idea how he looks. Her inability to remember his face does not interfere with other measures of attention in her brain. I ask her about my face right now. She’s looking at me, trying to gather as much information and memorable hints as possible. “Now... if I leave the room and come back, will you be able to identify me ?” Katy says she might, she now has an ample amount of data, but it will still take her quite some time. “It’s like these cats figures”, she points out to a picture of 3 cats (Figure 3):

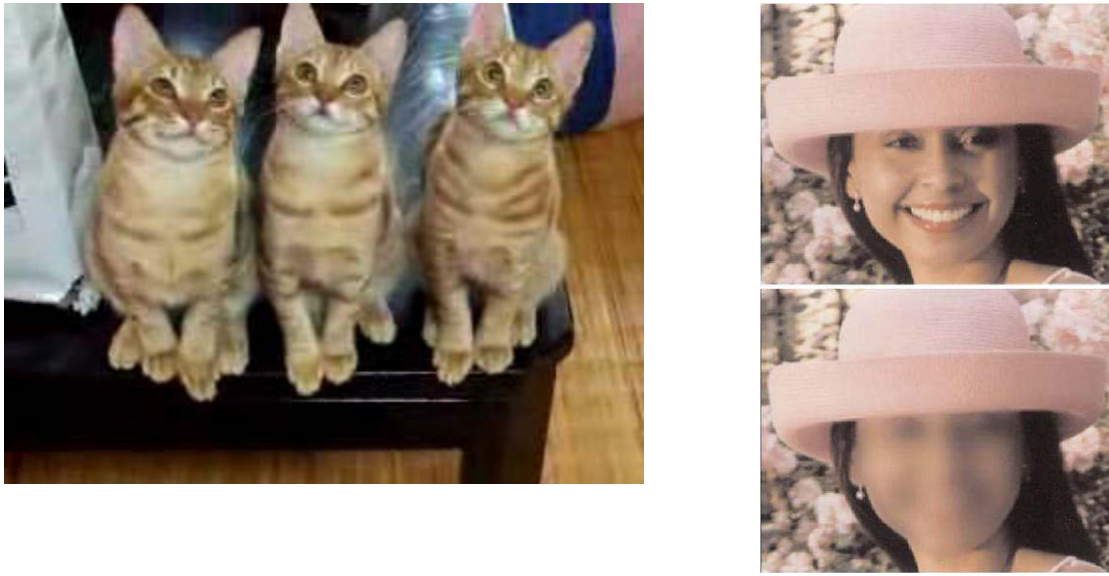


Figure 3 - Illustrations of the faces as seen by people with Prosopagnosia

Left panel. Would you be able to easily tell between these 3 cats without paying careful attention to details and trying to memorize them? This is similar to the way in which people with Prosopagnosia view faces. They can see the face, it just doesn't strike out as very different than other faces.

Right panel. Illustration of the way by which people with Prosopagnosia see a human face

"They look the same," I pointed out. "Exactly. That's how most people look to me. I can look carefully and find some features that will distinguish the cats from each other, but without careful attention to details, my brain doesn't immediately find the differences". I look at the cat and try to tell them apart. If one of them now entered the room, would I be able to identify it? surely not as quickly and easily as I would if Christof entered the room. Or my husband of 3 years...

“I once went to pick my husband from the airport. He went for a week on business travel. While there he cut his hair and changed his style a bit.... I circled LAX numerous times until I could actually identify him. And that’s someone I have seen a lot... Let’s put it this way. If you now talk to me, and while speaking you untie your hair tie and let your hair loose, which will change your face significantly, I will not be able to know it’s you anymore. I mean, consciously I will know that it’s obviously the same person talking to me, and I will see that it’s the same clothing and voice and everything, but perceptually – it will feel like one of those change blindness experiments.... I will just think that someone else replaced you in the middle of the sentence and it’s a practical joke. You will be a different person to me.”

Law and Order

I grilled Katy and Mike with questions about their life with no faces for hours, and the anecdotes and interesting stories can become a book one day. (Some of these are not appropriate and therefore should be kept out of this thesis). I want to add just one more story that I found of special interest. American people with Prosopagnosia are still American people, which means they watch TV a lot. However, they watch it differently. Katy and her husband really like the TV show “Law and Order”. I have never watched Law and Order in my life, but just from the name and my Wikipedia reading it is obvious to me that the show is about police and the judicial system. The program generally follows a two-part format, with the first portion of each episode devoted to the investigation of a crime and the second portion depicting its prosecution. Along this storyline, each part has its protagonist. Law and Order is noted for its revolving cast, but the current season portrayed the District Attorney Jack McCoy, played by Sam Waterston, and the police senior detective Kevin Bernard, played by Anthony Anderson

as lead characters. Jack McCoy is a man in his 50s. He has gray hair, and his face is somewhat wrinkled and serious. Kevin Bernard has a very short-cut hair, he has a small mustache, and he generally is much bigger than the slim McCoy. More than anything, Bernard is black and McCoy is white. The two are easily distinguishable. In fact, the only thing I could think of that the two share is that they are both men who repeatedly wear suits on the show. That's it.

Here are pictures of the two (Figure 4):



Figure 4 - District Attorney Jack McCoy and Detective Kevin Bernard

Nevertheless, for months Katy was following the series religiously as a show about a detective who arrests criminals and brings them to justice, standing before the judge and stating his case. The fact that two different teams are involved in each part of the episode, the fact that many storylines became confusing if you assume that the two are one, or the fact that it doesn't really make sense for a detective to also be the attorney didn't change a thing to Katy. She just thought "it's Hollywood" and "everything goes" in a TV series. It took months before, by accident, in a random conversation with her husband where she actually pointed out this absurdity in the series, that he actually realized her mistake. "Those are two different people," he said. Then Katy looked carefully and started noting the differences in hairstyle, body shape,

and skin color. “From then on it was very obvious to me that these are two people, but beforehand — since this was not my assumption — I just didn’t see the difference.”

Jay Leno

While it is somewhat reasonable for Katy to not pay careful attention to actors and gestures on TV, Mike is in fact a film writer in Hollywood. It’s his job to pay attention to these details and care for them. When I asked Mike about similar encounters with TV shows or films he told me that he repeatedly suffers from them. “You know those films where you follow the plot of the guy who is arguing throughout the entire film for his innocence and fiercely claims he did not do it, but then in the last scene, after he was lawfully found not guilty, is seen smiling to himself, and you now know he actually *did* do it? Well, I always miss this. My wife has to tell me what the facial expressions are and what they mean at the end of movies. Even ones I wrote.”

Finally, I asked Katy and Mike about the features they choose in a face, to tell people apart. Are those the most distinct features, or are they the ones that remain the same the most. I asked them for instance if the identifying features I attribute to a person would be the ones they choose. “Would it necessarily be Jay Leno’s chin, or do you find completely different features to identify Leno by?” The answer surprisingly was that the features used are almost never the ones we care about. “I just don’t notice his chin,” said Katy. “Tom Cruise’s smile is probably very unique, but to me his haircut is much more distinguishable,” added Mike. So what is it about the face that makes these people notice it, read its content, its emotional metrics and values, know it is there — yet, have this unique disorder where they just are not able to attend to it.



Figure 5 - Caricatures of Jay Leno and Tom Cruise

Face blindness

I would never have gotten to meet or interact with Mike and Katy if it wasn't for Doris Tsao. Doris was a junior faculty member when I first heard about her coming to Caltech. I attended her job talk beforehand and found her work relevant to mine and very impressive. Doris is only two years older than me but already has a notable work on face perception. Being the only female faculty that shares my interest in attention and a young enthusiastic scientist I felt she would become a great member of my thesis committee. I was worried, though. What happens if a young and enthusiastic new faculty, god forbid, would make me... work more? I was already at my fifth year of research when Doris joined Caltech and at the end of my thesis writing. The last thing I needed now was a "fresh" thinking that would end up delaying me by two more years. I came to offer Doris a spot on my committee with only one condition: "you are never to come up with ideas for new research for my thesis." Surprisingly, Doris said yes. More surprising was the fact that this was the first rule she broke.

I first discussed my thesis with Doris in her office at the beginning of January. We spent a couple of hours talking in detail about my work. I felt that she liked it. But then she dropped the bomb. "You know what will complete your story?" she said quietly. I reminded Doris of her promise... but, sadly, was tempted to hear. "What?". "Testing the same paradigms with

people with Prosopagnosia.” I was naïve. Instead of saying no right away and moving on, I underestimated Doris’ abilities. “But where would I get people with Prosopagnosia ? It’s probably impossible to find them.” “Well,” Doris suggested, “Let me try. If I succeed will you do it?” “Sure,” I answered, thinking it would be impossible for her to do so. Boy, was I misled. I didn’t know that Doris had internal information. She already knew Brad Duchaine from her time at Harvard. By the time I reached my lab (which is 5 minutes walk from her office), Doris had already emailed me and Brad, setting the introduction that would turn out to be one of the most fascinating parts of my research. Brad Duchaine is a scientist at Harvard and the University College of London. His research mainly deals with Prosopagnosia. Most notably, he maintains the <https://www.faceblind.org> website which is a forum for people who suspect they suffer from levels of Prosopagnosia. In the website one can participate in various simple test where one is asked to identify faces from objects, perform tests where one is asked to differentiate between similar faces, and identify faces in various conditions. The task is very easy for people who see faces, but nearly impossible for people with Prosopagnosia. According to a recent study by Duchaine’s graduate student at Harvard (not published yet, shown first as a personal communication at the Society for Cognitive Neuroscience conference in San Francisco, March 2009) face identification, like other traits such as I.Q. and language skills, is something that deteriorates with age, and peaks around age 20. So with due time most people can get to experience some extent of the face blindness conditions in Brad Duchaine’s website. The faceblind.org website is where both Katy and Mike started their journey towards understanding their unique disorder. Within 3 days after first talking to Doris about my work I started testing my two Prosopagnosia subjects. I am forever indebted to Doris for not accepting my one request.

Background

Most of the cases of Prosopagnosia that have been documented have been due to brain damage suffered after maturity from head trauma, stroke, and degenerative diseases (BC Duchaine, 2000; Kanwisher, 2000). These are examples of acquired Prosopagnosia: these individuals had normal face recognition abilities that were then impaired. This is the case with Mike, for instance. In contrast, in cases of developmental Prosopagnosia, the onset of Prosopagnosia occurred prior to developing normal face recognition abilities (adult levels of face recognition are reached during teenage years). Developmental Prosopagnosia has been used to refer to individuals whose Prosopagnosia is genetic in nature, individuals who experienced brain damage prior to experience with faces (prenatal brain damage or brain damage), and individuals who experienced brain damage or severe visual problems during childhood. However, these etiologies should be differentiated, because they are different paths to Prosopagnosia and so probably result in different types of impairment (BC Duchaine, Parker, & Nakayama, 2003; Galaburda & Duchaine, 2003; Kanwisher, 2000); they could be referred to as genetic Prosopagnosia, preexperiential Prosopagnosia, and postexperiential Prosopagnosia, respectively. In some cases, it may be difficult to determine the cause of Prosopagnosia, but many times individuals will either know that family members also suffer from Prosopagnosia or be aware of potential incidents that may have resulted in brain damage.

Individuals with developmental Prosopagnosia often do not realize that they are unable to recognize faces as well as others. Of course, they have never recognized faces normally so their impairment is not apparent to them. It is also difficult for them to notice, because individuals with normal face recognition rarely discuss their reliance on faces. As a result, there are a

number of individuals who have not recognized their Prosopagnosia until well into adulthood (B Duchaine & Nakayama, 2005).

There are a variety of explanations for Prosopagnosia. Most of these explanations propose that the procedures necessary for normal face recognition are not working properly (McKone, Kanwisher, & Duchaine, 2007). However, the explanations differ in their characterization of the impaired procedures. It appears that Prosopagnosia actually refers to a number of different types of impairments, so no one explanation will account for all cases of Prosopagnosia.

Relation to attention

One particularly interesting feature of Prosopagnosia is that it suggests both an attentional and unconscious aspect to face recognition. Experiments have shown that when presented with a mixture of familiar and unfamiliar faces, people with Prosopagnosia may be unable to successfully identify the people in the pictures, or even make a simple familiarity judgement ("this person seems familiar/unfamiliar"). However, when a measure of emotional response is taken (typically a measure of skin conductance) there tends to be an emotional response to familiar people even though no conscious recognition takes place (Bauer, 1984). This suggests emotion plays a significant role in face recognition, perhaps unsurprising when basic survival (particularly security) relies on identifying the people around you (BC Duchaine et al., 2003). It is thought that Capgras delusion (Ellis, Whitley, & Luaute, 1994; Förstl, Almeida, Owen, Burns, & Howard, 1991; Sacks, 1995; Tranel, Damasio, & Damasio, 1995) may be the reverse of Prosopagnosia. In this condition people report conscious recognition of people from faces, but show no emotional response, perhaps leading to the delusional belief that their relative or

spouse has been replaced by an impostor. In short – a unique and exciting disorder to study.
Definitely worthy of my time !

Epilepsy

Background

Epilepsy is a common chronic neurologic disorder consisting of recurrent, unprovoked seizures. Seizures are divided into two main types – generalized and partial. The generalized ones occur if the abnormal electrical activity affects all or most of the brain. These in turn are from these types: A tonic-clonic seizure is the most common type of generalized seizure. With this type of seizure your whole body stiffens, you lose consciousness, and then your body shakes (convulses) due to uncontrollable muscle contractions. Partial seizures are ones where the burst of electrical activity starts in, and stays in, one part of the brain. Therefore, you tend to have localized or “focal” symptoms. These tend to have signatures indicating the hemisphere in which they occur (i.e., moving of the right arm forward will act as a signature suggesting the focus is in the left hemisphere).

Epilepsy is caused by an underlying brain condition or brain damage. Some conditions are present at birth. Some conditions develop later in life. There are many such conditions. For example: a patch of scar tissue in a part of the brain, a head injury, stroke, cerebral palsy, some genetic syndromes, growths or tumors of the brain, previous infections of the brain such as meningitis, encephalitis, etc. The condition may “irritate” the surrounding brain cells and trigger seizures.

Epilepsy cannot be “cured” with medication. However, various medicines can prevent seizures. They work by stabilizing the electrical activity of the brain.

About 50 million people worldwide have epilepsy, making it a disease as common as diabetes. A list of notable people with epilepsy shows that many world leaders, scientist, and artists were able to live a healthy, productive life with epilepsy. In fact, nowadays the only immediate concerns rising from having epilepsy is the inability of people with the disease to be licensed to drive a car, and the chance of hurting themselves during a seizure. If those are controlled, life with recurring seizures can be perfectly normal. However, for some reason the “marketing” of diabetes is much better and it is much more spoken-of than epilepsy.

The word epilepsy is derived from the Greek *epilepsia*, which in turn can be broken into *epi-* (-upon) and *lepsis* (-to take hold of, or seizure) (Harper, 2001). In the past, epilepsy was associated with religious experiences and even demonic possession. In ancient times, epilepsy was known as the "Sacred Disease" because people thought that epileptic seizures were a form of attack by demons, or that the visions experienced by persons with epilepsy were sent by the gods. Epilepsy was often understood as an attack by an evil spirit, but the affected person could become revered as a shaman through these otherworldly experiences. However, in most cultures, persons with epilepsy have been stigmatized, shunned, or even imprisoned; in the Salpêtrière, the birthplace of modern neurology, Jean-Martin Charcot found people with epilepsy side-by-side with the mentally retarded, those with chronic syphilis, and the criminally insane. In Tanzania to this day, as with other parts of Africa, epilepsy is associated with possession by evil spirits, witchcraft, or poisoning and is believed by many to be contagious. In ancient Rome, epilepsy was known as the *Morbus Comitialis* (“disease of the assembly hall”) and was seen as a curse from the gods.

Surgery

Epilepsy surgery is an option for patients whose seizures remain resistant to treatment with anticonvulsant medications who also have symptomatic localization-related epilepsy – a focal abnormality that can be located and therefore removed. The goal for these procedures is total control of epileptic seizures, although anticonvulsant medications may still be required.

The evaluation for epilepsy surgery is designed to locate the "epileptic focus" (the location of the epileptic abnormality) and to determine if resective surgery will affect normal brain function.

Epilepsy surgery for focal seizures began more than a century ago and progressed with the technical innovations of EEG and neuroimaging. In the 1860s and 1870s, the pioneering clinical work of the epileptologist John Hughlings Jackson lay the groundwork for understanding the cortical localization of focal epilepsies, while the animal experiments of the neurophysiologists Gustav Theodor Fritsch, Eduard Hitzig, and David Ferrier gave parallel confirmation of Jackson's conclusions (Rosenow & Luders, 2001).

In the 1870s and 1880s, the first documented resections for epilepsy were performed, making use of the new cortical localization principles (Rosenow & Luders, 2001). In 1879 the surgeon William Macewen correctly localized and resected a frontal mass in an epileptic patient in Glasgow; the patient survived removal of the meningioma and his seizure disorder was cured (Rosenow & Luders, 2001). In 1884 the neurologists Jackson and Alexander Hughes Bennett localized a putative mass in another epileptic patient; the surgeon Rickman Godlee successfully resected the underlying tumor found at their hypothesized site (Bennett & Godlee, 1884).

In 1886 Victor Horsley publicly presented his guidelines for antiseptic and hemostatic brain surgery in humans, focusing on three cases of young men subject to “fits.” His lecture consisted of a detailed description of three epilepsy surgery cases. Jackson, on whose patient Horsley had operated, was in the audience and advised the conference attendees that surgery should be performed for seizures even in the absence of a mass lesion. Jackson's comments were paraphrased in the 1886 publication of Horsley's lecture: “Believing that the starting point of the fit was a sign to us of the seat of the ‘discharging lesion,’ he would advise cutting out that lesion, whether it was produced by tumour or not” (Engel, 1987). Jackson was in fact recommending the resection of what was later to be called “the epileptogenic zone,” regardless of the presence of structural abnormality.

The recognition that certain epileptic syndromes are “surgically remediable” has improved patient selection for epilepsy surgery. Surgically remediable epileptic syndromes are conceptualized as “conditions with a known pathophysiology and natural history that have a poor prognosis with purely medical treatment, but that respond well to surgical treatment” (Engel, 1987).

Surgical procedures

After presurgical evaluation has identified the epileptogenic zone, the operative procedure most likely to benefit the patient is planned. Epilepsy surgery includes both resective and disconnective procedures. Resective epilepsy surgery is comprised primarily of (a) anteromesial temporal resections and (b) neocortical resections, including both lesionectomy and resection of electrographically abnormal regions that have no structural correlate visible on MRI.

Disconnective procedures include (c) multiple subpial transection, (d) corpus callosotomy, and (e) functional hemispherectomy.

The majority of resective procedures performed for epilepsy are variants of anteromesial temporal resections, since mesial temporal lobe epilepsy is both the most common and the most medically refractory focal epilepsy (Semah, 1998). In addition, a disconnective procedure for temporal lobe epilepsy referred to as temporal lobectomy has recently been reported in a small case series with preliminary outcome data (J. Smith, VanderGriff, & Fountas, 2004).

Anteromesial temporal resections. Anterior temporal lobectomies can be divided into (1) standardized anatomic resections of the anterior temporal lobe, and (2) tailored resections. At most epilepsy surgery centers, standardized anterior temporal lobectomy is performed when concordant data gathered during presurgical evaluation supports the diagnosis of mesial temporal lobe epilepsy associated with hippocampal sclerosis (Fried, 1993). Tailored resections are typically performed in those temporal lobe epilepsy cases that do not fit the stringent criteria for mesial temporal lobe epilepsy associated with hippocampal sclerosis and in those cases in which the epileptogenic zone is in dangerous proximity to eloquent cortex (Ojemann & Silbergeld, 1993).

An important note that I feel is important to mention in this context is the possibility of this type of work having a limited “window of opportunity”. There’s a chance that this type of research that is based on invasive surgery in these epilepsy patients’ brain will end in the coming future. This is due to the advances in medications and treatment possibilities for patients with epilepsy that potentially could lead to better non-invasive treatments for this

disorder that does not require any surgery. This of course will end this type of research that is fully dependant on the clinicial need for this surgery.

Electrode placement

Electrodes are placed stereotactically with MR imaging and angiographic guidance. Before surgery each patient undergoes placement of a stereotactic headframe, and then a detailed MR image is obtained using a spoiled-gradient sequence, followed by cerebral angiography. Both MR and digital subtraction (DS) angiography images are transmitted to a workstation in the operating room, and surgical planning is then performed, with selection of appropriate temporal and extratemporal targets and appropriate trajectories based on clinical criteria. Special attention is given to the venous phase of the angiogram so that the cortical surface veins can be avoided at the chosen trajectories. A dynamic multi-image environment permits simultaneous display of MR and DS angiography images, with rapid viewing of potential trajectories in different planes before the final selection is made. The patient is then taken to the operating room and general anesthesia is induced, after which multiple electrodes are placed, usually bilaterally and orthogonally from lateral to medial. At each entry point a twist-drill hole is made, the dura is coagulated and punctured, and a screw guide is inserted into the bone. The electrode is then introduced with a stylet through the guide screw to the correct depth, the stylet is withdrawn, and the microwires with the microdialysis probe are introduced through the lumen of the electrode. Up to 9 microwires (40 μm diameter) are inserted through each lumen (“macro electrode”). The macro-electrodes contain 6-7 contacts approximately 1.5 mm wide with separations of 1.5-4 mm (see Figure 7).



Figure 6 - A photograph showing one of the patients

The patient is seen with the electrodes implanted while he was in the ward. Taken from (Kreiman, 2001). The electrodes are implanted in his brain in order to monitor the focus or foci of his seizures. Permission from the patient was obtained prior to using the picture.

From these contacts, continuous EEG data are acquired 24 hours a day. The microwires are composed of foamvar-insulated platinum / 20% iridium and typically have an impedance in the range of 0.2 to 1 M Ω .

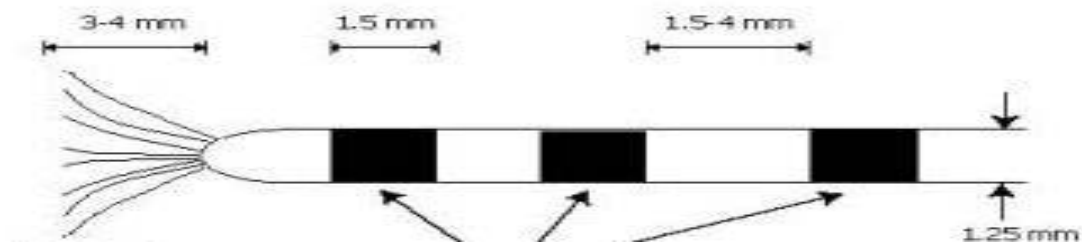


Figure 7 - Electrodes implanted in patient's brain

A schematic description of the type of electrodes most often used in temporal lobe targets. Platinum/iridium contacts of approximately 1.5 mm length along the electrode were used to acquire clinical wide band EEG data. Through the lumen of the 1.25 mm diameter electrodes, 8 or 9 platinum/iridium microwires were inserted. Macro-electrodes were fabricated at Ad-Tech (Racine, WI). Micro-electrodes were modified at UCLA. Microwires extended 1 to 3 mm from the tip of the electrode.

The separation between the tip of the microwires can range up to 4 mm, and their spread inside the brain is not controlled by the surgeon. Finally, a cap is secured over the guide screw to prevent cerebrospinal fluid leakage. This procedure is repeated for each macro-electrode. Following electrode placement the patients are monitored in a special unit on the neurosurgical ward for a period of 1 to 3 weeks, until a sufficient number of spontaneous seizures had been recorded. Typically, this takes approximately seven to ten days. The first one or two days after surgery, patients are under different types of medication and are not up to doing any tests; all the tests are typically performed on days two through seven after the implantation of the electrodes.

Prior to removing the electrodes, MR images are obtained to confirm their location (see Figure 8). The patient is then given a local anesthetic and the electrodes and guide screws are removed.

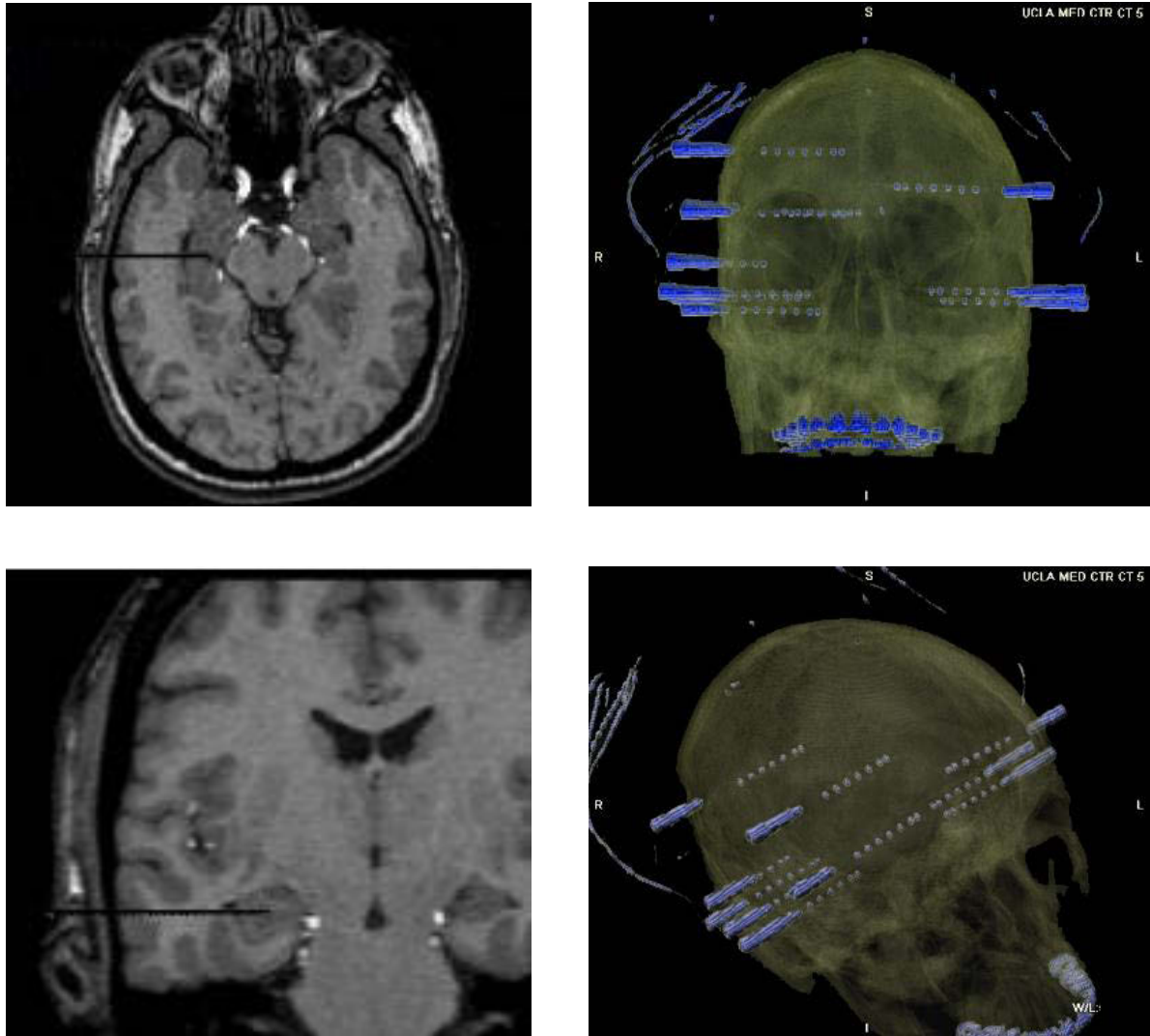


Figure 8 – MRI image of one of the patients

The position of the electrodes was determined by obtaining a structural magnetic resonance image. Here we show an example of an MRI obtained in one of the patients (**left**) indicating the position of one of the probes in the right hippocampus. The **top-left** image shows an axial

image and the **bottom-left** image shows a detail of the coronal section. MRI scans were obtained during the week of chronic monitoring with the electrodes implanted; post-operative CT and MRI were also obtained. The CT was co-registered with the MRI structural information for anatomical verification. The distal end of the electrode can be seen here in the images and included the platinum/iridium microwires schematized in the previous figure. The MRI was obtained at 1.5 Tesla. This allowed us to indicate the anatomical region of the electrode but not to specify which layer or specific sub-region within each area we recorded from. On the right are two corresponding structural illustrations of the electrodes implanted.

While the neurosurgeon typically placed between 11-16 macro-electrodes in the patients' brains for localization of the seizure, we always used only 8 of them for each recording due to a limitation of our acquisition system that allows the usage of only 8 x 8 microelectrodes. For each session we therefore selected the 8 electrodes from which we wished to test (the "montage"). While some patients had electrodes in various areas such as the pre-supplementary motor area, anterior and posterior cingulate, etc., we typically used the entire medial temporal lobe montage that included the hippocampus, parahippocampal gyrus, entorhinal cortex, and amygdala from each hemisphere – giving the required 8 electrodes. Since in some sessions we did not see neurons at the tip of some of these electrodes (due to signal problems, impedance issues, etc.) we collected some data from other areas.

Figure 9 shows the general scheme of regions we recorded from in a typical montage.

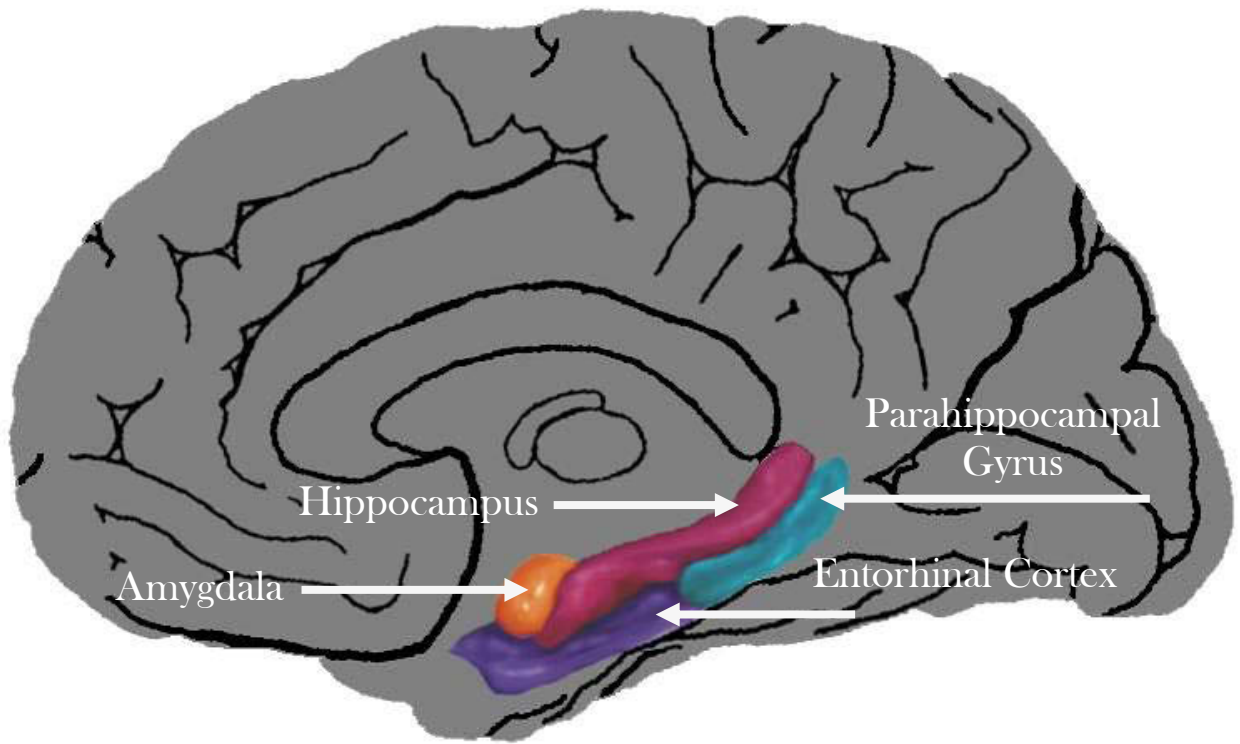


Figure 9 - Regions we recorded from in the medial temporal lobe



Attention and Competition

*If our brains were simple enough
for us to understand them, we
would be too simple to understand
them.*

Ken Hill

Introduction

Competition is defined as rivalry between individuals for allocation of resources. It arises whenever two or more parties strive for limited resources. Competition occurs naturally between living organisms which co-exist in the same environment. We compete for water, food, mates, wealth, prestige, and fame. While psychology defines competition every so often as a mechanism that can yield improvement if the product of the competition is targeted at making a better output, most often the associations that rise from competition are negative.

The Latin root for the verb “to compete” is “competere”, which means “to seek together” or “to strive together”³. So which is it? To work WITH an entity or to work AGAINST one? For the sake of this work we will regard competition as a process by which two entities try to use the same resource while only one of the two can make use of it. The processes will be neurons, and the resource will be our conscious percept.

There can be only one

The end of my thesis writing era was marked by the election of Barack Obama as president of the United States. After his November 4th election there were two months where he was president-elect while George W. Bush was still the president. Numerous times when major events occurred in the world arena Obama was asked to comment on those. These were a war that broke out in Israel, a political crisis with Russia, and various economic meltdowns that marked the beginning of an economic recession. The president-elect repeatedly refused to

³ dictionary.com

comment on any of the events, stating repeatedly the same mantra: “There’s only one president at a time.”⁴ This suggested that according to Obama, until his official inauguration January 20th, the entire stage was governed by his predecessor. Only one president at a time.

In my early 20s when I was taking my first flying lessons on a small Cessna-152 at the Herzeliya airport in Israel I always wondered how the air traffic control manages to monitor so many aircraft taking off and landing at the small airport simultaneously. I once asked for permission and went to visit the control tower, and was stunned to see that a single person was operating the entire fleet of dozens of aircraft entering and leaving the airfield. I – naïvely – asked the controller how he could land so many airplanes in this airfield with such an intense schedule. His answer remained with me and made sense towards my writing along these lines: “It’s simple. You just do it one airplane at a time.”

In the classic movie “Highlander” the premise is a continuous combat between the warriors trying to win “The Prize” (that is eventually revealed to be mortality). Along the plot, through the continuous fights between the warriors, one repeatedly hears the same shout upon the decapitation of one of the immortals pursuing the prize: “There can be only one.”

At any given moment we are aware of a single thing. We are conscious of one aspect of the environment. At the same time, a sequence and an ample amount of signals from the environment penetrate our brain – sounds, images, smells, sensory inputs, and internal signals are all simultaneously sent to the brain. Without a serial queue, or a central processing unit there’s a need for an officer, a cop, a signaling system that will determine which entity is

⁴ <http://www.politico.com/news/stories/0109/17090.html>

actually the one perceived and which information is to be processed later or in the background. Which is to be sent to a different module of processing and thus be evaluated simultaneously and have its output be available to the higher mechanisms, and how to determine who is the current president, or airplane that we are in fact thinking of at a given moment. This process is commonly looked at as a spotlight on a stage lighting one dancer out of many at any given moment. This spotlight is referred to as our attention.

Attention

A typical visual scene contains many different objects, not all of which can be fully processed by the visual system at any given time. Thus, attentional mechanisms are needed to limit processing to items that are currently relevant to behavior (Broadbent, 1958; Bundesen, 1990; Desimone & Duncan, 1995; Neisser, 1967; Treisman, 1969; J. Tsotsos, 1990). Probably the dominant neurobiological hypothesis to account for attentional selection is that attention serves to enhance the responses of neurons representing stimuli at a single behaviorally relevant location in the visual field (Colby, Duhamel, & Goldberg, 1996). This enhancement model is closely related to older “spotlight of attention” models in psychology, in which visual attention serves to limit processing to a single locus of variable size in the visual field. According to this classical view, a behaviorally relevant object in a cluttered field is found by rapidly shifting the spotlight from one object in the scene to the next, until the sought-for object is found. Attention essentially serves as an internal eye that can shift its focus from one location to another. Because all visual attention is inherently spatial according to this view, even objects defined by their shape or color must be found by examining candidate objects with the serially scanning spotlight, unless the object is so distinctive that it “pops out” and automatically attracts the attentional spotlight. The neurobiological spotlight hypothesis has the advantage of both simplicity and a clear relation to the neuronal enhancement effects seen in the oculomotor system and throughout most of the visual cortex for stimuli that are the targets of eye movements (Desimone, 1998).

Competition

Competition is a process, not just a result. With important exceptions, most theories of competition concern what is left when competition is over. The alternative is to start with the process of competition and work toward its results. This is the route we will use in this work.

Suppose that you are looking for a face in a crowd. Two basic phenomena occur while processing that scene. First, not all faces can be processed at the same time, that is, there is limited processing capacity. Second, while processing a particular face, one is able to filter out the unwanted information in the scene, that is, there is selectivity. The biased competition theory of selective attention rests on three general principles that conceptualize these basic observations further (Duncan, 1985). First, of the many brain systems that represent visual information (sensory and motor, cortical and subcortical), most are competitive. Within each system, a gain of representation for a particular visual object will be at the expense of other objects' representations. Such competitive interactions among multiple objects (such as the faces in a crowd) occur automatically and operate in parallel across the visual field. Second, competition is controlled within and across brain systems. If one looks for a particular object (e.g., a friend's face), units matching the internal "template" of that object will be pre-activated and therefore gain an advantage by receiving an increased processing weight. Thus, such top-down mechanisms introduce bias signals that help resolve the ongoing competition. The competition among multiple objects can also be biased by bottom-up mechanisms that separate figures from their background, or constitute objects by principles of perceptual organization. And third, the competition between systems is integrated. As a visual object gains dominance in representation within one system (e.g., visual cortex), it will tend to gain similar dominance

in other systems (e.g., higher-order frontal and parietal areas). An example is given by representations of visual space. All units that represent a certain location in multiple spatial maps will be activated together, when the object at that location gains dominance in the system.

Behavioral data in neuroscience

Many studies directly address the nature of competition for attention in the human brain. In one simple type of experiment, two objects are presented in the visual field. Subjects must identify some property of both objects, with a separate response for each. Such studies reveal several important facts.

First, dividing attention between two objects almost always results in poorer performance than focusing attention on one. Identifying simple properties of each object such as size, brightness, orientation, or spatial position gives much the same result as identifying more complex properties such as shape (Duncan, 1984, 1985, 1993). A possible exception is simple detection of simultaneous energy onsets or offsets (Bonnell, Stein, & Bertucci, 1992). Second, as long as the experiment uses brief stimulus exposures and measures the accuracy of stimulus identification, the major performance limitation appears to occur at stimulus input rather than subsequent short-term storage and response. For example, interference from processing two objects is abolished if they are shown one after the other, with an interval of perhaps a second between them (Duncan, 1980), even though the two responses called for must still be remembered and made together at the end of the trial. Third, interference is independent of eye movements. Even though gaze is always maintained at fixation, it is easier to identify one object in the periphery than two. Fourth, interference is largely independent of the spatial separation between two objects, at least when the field is otherwise empty (Sagi & Julesz, 1985;

Vecera & Farah, 1994). Though attention is sometimes seen as a mental spotlight illuminating or selecting information from a restricted region of visual space (Eriksen & Hoffman, 1973; Posner, Snyder, & Davidson, 1980), performance seems not to depend on the absolute spatial distribution of information. An enduring issue is the underlying reason for between-object competition. It has often been argued that full visual analysis of every object in a scene would be impossibly complex (Broadbent, 1958; J. Tsotsos, 1990). Competition reflects a limit on visual identification capacity. Equally strong, however, has been the view that competition concerns control of response systems (Allport, 1980; Deutsch & Deutsch, 1963). Certainly, some response activation often occurs from objects a person has been told to ignore (Eriksen & Hoffman, 1973), which shows that unwanted information is not entirely filtered out in early vision. Very probably, competition between objects occurs at multiple levels between sensory input and motor output (Allport, 1980).

Neural basis for competition

If the nervous system had unlimited capacity to process information in parallel throughout the visual field, competition between objects would presumably be necessary only at final motor output stages. Before discussing these motor stages, we first consider what limitations in the visual system make competition necessary at the input. Objects in the visual field compete for processing within a network of visual areas (Felleman & Van Essen, 1991). These areas appear to be organized within two major corticocortical processing pathways, or streams, each of which begins with the primary visual cortex, or V1 (see Figure 10).

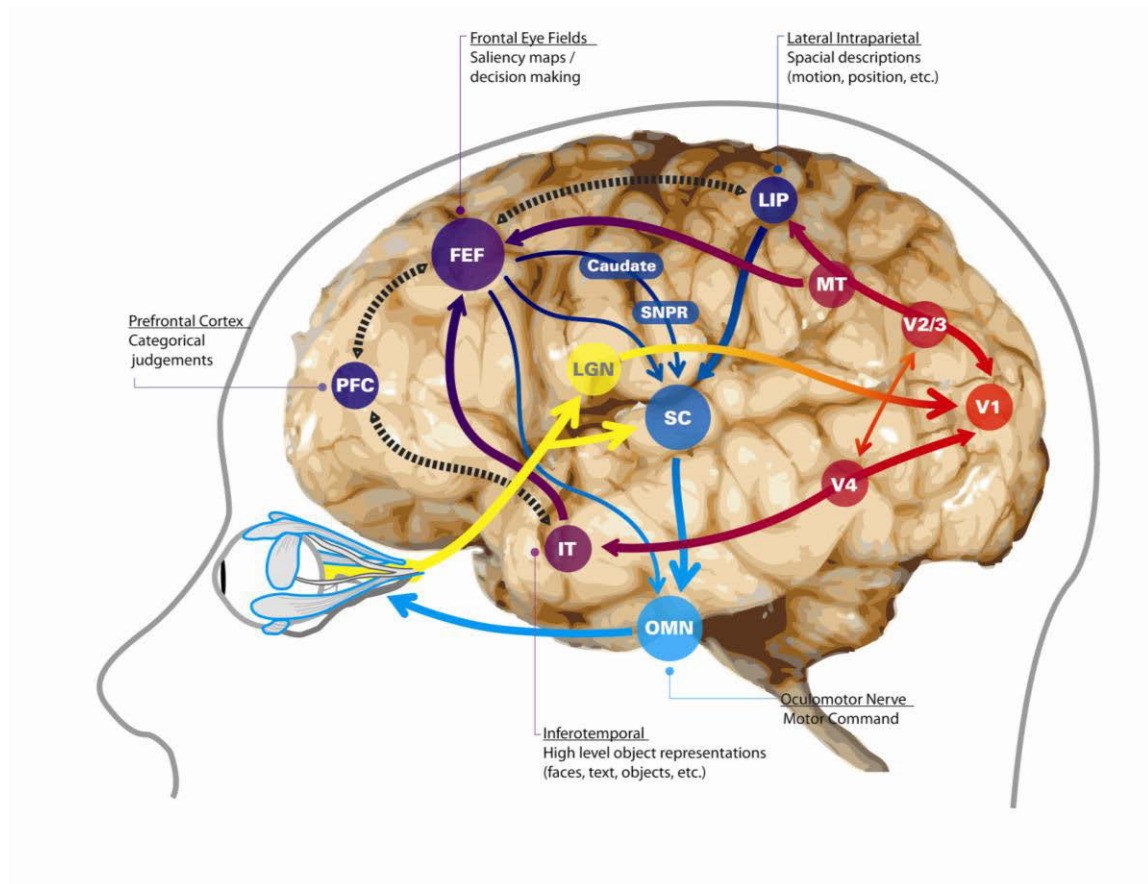


Figure 10 - Processing pathways in the visual stream

Striate cortex, or v1, is the source of two cortical visual streams. A dorsal stream is directed into the posterior parietal cortex and underlies spatial perception and visuomotor performance. A ventral stream is directed into the inferior temporal cortex and underlies object recognition. Both streams have further projections into the prefrontal cortex. Adapted from (Mishkin, Ungerleider, & Macko, 2001; Wilson, Scalaidhe, & Goldman-Rakic, 1993). For a “wiring diagram” of the areas and connections of the two streams, see (Felleman & Van Essen, 1991).

The first, a ventral stream, is directed into the inferior temporal cortex and is important for object recognition, while the other, a dorsal stream, is directed into the posterior parietal cortex and is important for spatial perception and visuomotor performance (Mishkin et al., 2001;

Ungerleider & Haxby, 1994). Since competition impacts object recognition, we would expect to find one basis for it in the ventral stream. The ventral stream includes specific anatomical subregions of area V2 (thin and interstripe regions), area V4, and areas TEO and TEi in the inferior temporal (IT) cortex. As one proceeds from one area to the next along this pathway, neuronal properties change in two obvious ways. First, the complexity of visual processing increases. For example, whereas many V1 cells function essentially as local spatiotemporal energy filters, V2 neurons may respond to virtual or illusory contours in certain figures and IT neurons respond selectively to global or overall object features, such as shape. Second, the receptive field size of individual neurons increases at each stage. As one moves from V1 to V4 to TEO to TE, typical receptive fields in the central field representation are on the order of 0.2, 3, 6, and 25° in size, respectively. Large receptive fields may contribute towards the recognition of objects over retinal translation. These receptive fields can be viewed as a critical visual processing resource, for which objects in the visual field must compete. If one were to add ever-more-independent objects to a V4 or IT receptive field, the information available about any one of them would certainly decrease. If, for example, a color-sensitive IT neuron were to integrate wavelength over its large receptive field, one might not be able to tell from that cell alone if a given level of response was due to, say, one red object or two yellow ones or three green ones at different locations in the field. Such ambiguity may be responsible for the interference effects found in divided attention. This ambiguity may be reduced, in part, by linking objects and their features to retinal locations. It is sometimes presumed that location information is absent from the ventral "what" stream altogether and must be supplied by the dorsal "where" stream. In fact, the ventral stream itself contains information about the retinal location of complex object features (Hung, Kreiman, Poggio, & DiCarlo, 2005). V4 and TEO

neurons process relatively sophisticated information about object shape and have retinotopically organized receptive fields. At any given retinotopic locus in these areas, receptive fields show considerable scatter. One could, in principle, derive information about the relative locations of nearby features from a population of cells with partially overlapping fields the same way one could derive information about a specific color from a population of neurons with broad but different color tuning. Similarly, although receptive fields in the IT cortex may span 20–30 degrees or more, they are not homogeneous. Typically, the fields have a hot spot at the center of gaze, which may extend asymmetrically into the upper or lower contralateral visual field. Although the stimulus preferences of IT neurons remain the same over large retinal regions, for a large minority of cells the absolute response to a given stimulus changes significantly with retinal location, i.e., cells are tuned to retinal location the same way they are tuned to other object features. Thus, in principle, objects and their locations might be linked to some extent within the ventral stream. Even so, parallel processing across the visual field is likely to be limited.

To sum up, retinal location, as with other object features, is coarsely coded in the ventral stream. Information about more than one object may, to some extent, be processed in parallel, but the information available about any given object will decline as more and more objects are added to receptive fields. Therefore, objects must compete for processing in the ventral stream, and the visual system should use any information it has about relevant objects to bias the competition in their favor. This issue which we term selectivity, is considered in later sections. If the dorsal stream receives its visual input in parallel to the ventral stream as the anatomy suggests (Desimone & Ungerleider, 1989), then it is presumably faced with competition among objects as well. As in the IT cortex, receptive fields in the posterior parietal

cortex are very large, and it seems likely that increasing the number of independent objects in the visual field will eventually exceed the capacity of the parietal cortex to extract the locations of each of them in parallel. Likewise, neural systems for visuomotor control must also deal with competition, to the extent that distractors are not already filtered out of the visual input (Munoz & Wurtz, 1993). Ultimately, for example, it is possible to move the eyes to only one target at a time. A critical issue is how selectivity is coordinated across the different systems so that the same target object is selected for perceptual and spatial analysis as well as for motor control.

Models of attention as competition between brain resources on the conscious percept

Multiple stimuli compete for neural representation in visual cortex

The first prediction of a competition theory is that objects compete for neural representation in visual cortex. Evidence from both single-cell physiology in monkeys and neuroimaging suggests that multiple stimuli present at the same time within a neuron's receptive field are not processed independently, but interact with each other in a mutually suppressive way. In a study discussed in chapter 20 we will show evidence that this competitive interaction follows past the receptive field to areas in the medial temporal lobe. In single-cell studies in monkeys, responses to a visual stimulus presented alone in a receptive field were compared to the responses evoked by that stimulus when a second one was presented simultaneously within the same receptive field. The responses to the paired stimuli were found to be smaller than the sum of the responses evoked by each stimulus individually and turned out to be a weighted average of the individual responses (Reynolds, Chelazzi, & Desimone, 1999). These suppressive interactions among multiple stimuli present simultaneously demonstrate the idea that these stimuli are competing for representation by single neurons in the visual cortex. In the human brain, evidence for neural competition has been found using an fMRI paradigm (Beck & Kastner, 2005, 2007; Kastner, De Weerd, Desimone, & Ungerleider, 1998; Kastner et al., 2001).

Competition can be biased by top-down and bottom-up mechanisms

Competition can be biased via top-down control mechanisms or via bottom-up stimulus-driven mechanisms. The term top-down is used to refer to biases that are generated by the cognitive demands of the task, and not by the competing stimuli themselves. Top-down biases are thought to exert influence on the visual cortex, at least initially, via feedback mechanisms from the frontoparietal cortex. The top-down mechanism that I will focus on is spatially directed attention to a location or feature of a stimulus. However, one should keep in mind that other top-down mechanisms related to memory processes, or emotional and motivational behavior, to name just a few, can introduce top-down biases as well. Again, both single-cell physiology and fMRI evidence supports the second tenet of biased competition theory. Single-cell recordings with monkeys has shown that attention can be used to filter unwanted information during a competition by modulating competing stimuli (Beck & Kastner, 2008). Attention may resolve the competition among multiple stimuli by counteracting the suppressive influences of nearby stimuli, thereby enhancing information processing at the attended location. This may be an important mechanism by which attention filters out information from nearby distracters. Studies with humans suggest that areas at intermediate levels of visual processing such as V4 and the inferotemporal area are important sites for the filtering of unwanted information by counteracting competitive interactions among stimuli at the level of the receptive field.

When attention is studied almost independent of a stimulus, when animals or humans are cued to attend covertly to a location within the receptive field before a stimulus is presented (Luck, Chelazzi, Hillyard, & Desimone, 1997) the baseline firing rate of neurons is increased. In the framework of biased competition theory, baseline increases can be thought of as a direct

measure of the increased processing weight that a location receives during the allocation of attention.

Competition cannot only be resolved by top-down mechanisms such as spatially directed attention, but also by bottom-up stimulus-driven signals. Unlike top-down biases, bottom-up biases have their source in the visual stimulus itself. For instance, competition may be biased in favor of a visual salient item that contrasts with its background. Many bottom-up factors are likely to be generated in the visual cortex itself. However, this is not necessarily the case. For instance, processing may be biased in favor of an emotionally salient item via connections with the amygdala. The critical aspect of a bottom-up bias is that it is something about the stimulus itself that induces the bias, as opposed to being imposed by the goals of the observer. Thus, bottom-up biases will affect processing in the visual cortex even in the absence of top-down mechanisms directed to the stimuli in question. Pop-out experiments, in which a single item differs from the others in color, orientation, size, etc, have shown that bottom-up based competition indeed alters the attention paid to the unique stimulus. Bias competition theory also predicts other bottom-up effects of the stimulus on competition that may not result in a bias *per se*. More specifically, it predicts that competitive interactions should be modulated by perceptual grouping. As originally noted by the Gestalt psychologists (Rubin, 1958), visual stimuli in cluttered scenes may be perceptually grouped according to their similarity, proximity, common fate and other stimulus properties, linking elements of a scene that are likely to belong together and thereby segmenting the scene into a more limited number of object-based perceptual units. Desimone and Duncan (Desimone & Duncan, 1995) predicted that competition should occur between these perceptual groups, but not among multiple items within a perceptual group. While this hypothesis was contradicted by a few recent studies, it is

still a governing claim that there is less competition within a group of similar entities in a search task, for instance, than across groups. The overall principle suggested by Duncan (Duncan, 1996) that competition occurs between objects, and not features within that object considers grouping as one of a number of segmentation processes that define an object, and therefore competition will occur among members of a perceptual group in a bottom-up fashion.

While commonly top-down and bottom-up influences on competition are discussed separately because they are believed to be dependent on qualitatively different mechanisms, these biases are likely to interact in everyday life. We here show a quantitative method (chapters 10 – 16, 20 – 21) to distinguish their exact interferences and weighted contribution in a unique single-cell recording from human brains. The third tenet of biased competition theory which supposes that selection of a target object emerges through the integration of many separate competitive systems, such that when an object gains dominance in one system (e.g., visual cortex), it will tend to gain similar dominance in other systems (e.g., higher-order frontal and parietal areas), is the least supported by empirical evidence. However, an example for the integration principle has recently been found in the spatial domain. Attention effects in the visual cortex have been shown to be highly spatially specific. For example, directing attention to a particular region of space enhances responses only in cortical areas with a representation of the attended location (Brefczynski & DeYoe, 1999).

In summary, biased competition theory of selective attention, as proposed by Desimone and Duncan (Desimone & Duncan, 1995) has had a tremendous influence on the field of visual attention. Moreover, evidence now exists in favor of all of its most basic principles. However, studies with directed recordings from brains of humans, and tasks that directly exhibit

competition between stimuli, brain areas, and top- versus bottom-up attention mechanisms are, to the best of my knowledge, nonexistent and therefore this work is the first in providing a “spotlight” of attention on that particular aspects of the question.



About Face

To a true artist only that face is beautiful which, quite apart from its exterior, shines with the truth within the soul

Mahatma Gandhi

Faces are one of the most important cues for us. We see them everywhere and all the time. When we talk to people we commonly look at their faces. When we judge people's attraction, first and foremost we refer to their looks, and by looks we mainly refer to how visually appealing their face is (Jonh Sirl, personal communication). We remember people by their face. We put faces as a signature of people's existence on our passport, drivers license, or any other mean of unique identification. We want to know people's "true face" when we ask who they really are. We buy tickets at "face value". People's relationship to god is commonly exhibited by them seeing faces in places where god intervenes in our lives, or should have intervened. During the September 11 attack on the twin towers, people saw the "face of god" in the smoke coming out of the burning buildings. An avid believer saw the face of The Virgin Mary on a piece of toast she later sold for thousands of dollars on eBay (see Figure 11). Faces have various proverbs and compound words used in reference to them in general slang, and are also the subject of various popular icons (see boxes in this chapter). Our study of faces was geared towards combining these vastly powerful attractors and important cues in various models of attention, so as to provide evidence of the importance of emotionally salient and social cues in the attraction of attention. This introduction will shed light on some of the common studies of faces to date.

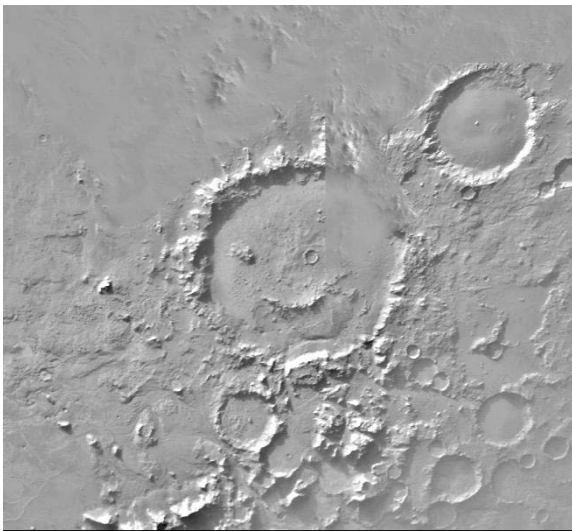
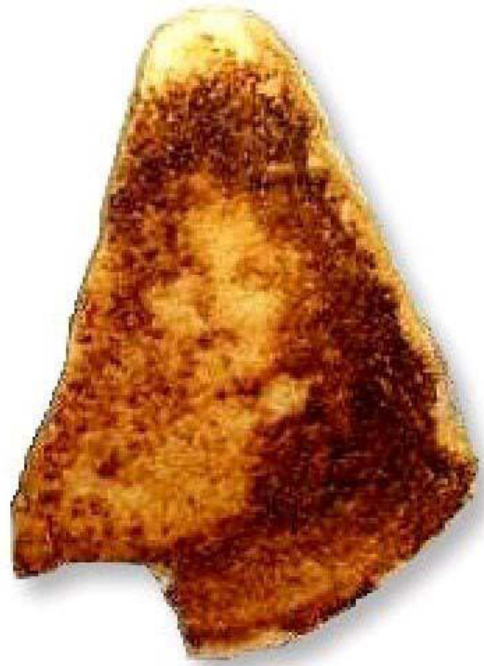


Figure 11 - People see faces everywhere.

Top left. The face of the devil seen by people in the smoke coming out of the twin buildings during September 11, 2001

Top right. The face of Maria, Jesus' mother on a piece of toast which was later sold on eBay for \$22,000

Bottom left. Smiley face seen on the surface of the moon

Bottom right. A sign of life on Mars greeting humans in the first images captured through the Hubble telescope

Introduction

My interest in faces as a cue affecting attention started when Christof first referred me to a study conducted by Hochstein and Hershler suggesting that faces show a pop-out effect in attention tasks (Hershler & Hochstein, 2005). As discussed in details in the attention and competition introductory chapter, pop-out effects on attention are commonly used to study attention as a bottom-up driven mechanism. The study by Hochstein showed that faces, like orientation, color, hue, and motion have a bottom-up driven component that makes observers attend to faces rapidly when exposed to a grid of stimuli, one of which is a face. Their results, tested in various grid sizes using pure reaction-time measures clearly indicate that people find faces faster than they find any other competing entity (e.g., cars). See Figure 12 for an illustration of the experiment and effect.

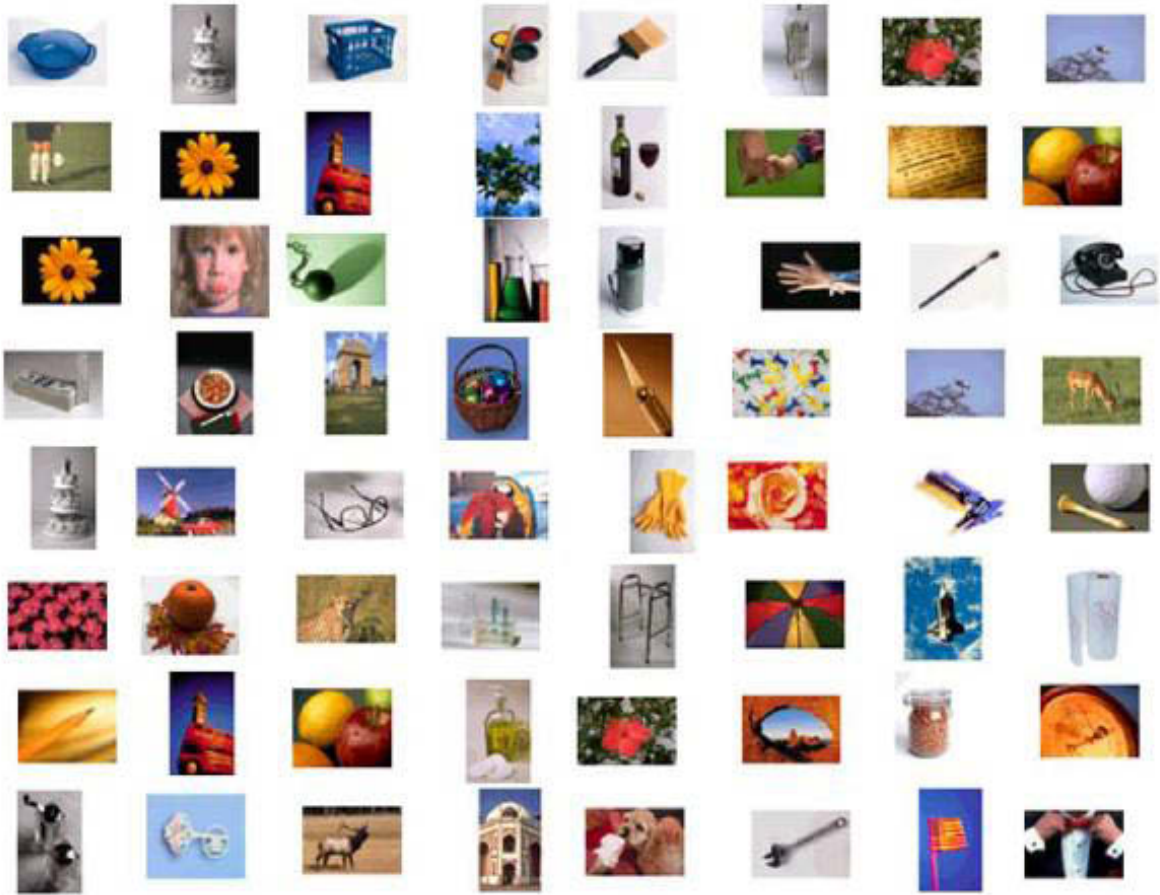


Figure 12 - Example of a face pop-out effect

In this 8 x 8 grid of photographs. Note that the picture of a face pops out from among the other, distractor photographs. Image taken from (Hershler & Hochstein, 2005).

The results by Hershler & Hochstein triggered a controversy between the authors who suggested in their paper the conclusion that face search may reflect high-level activity with generalization over spatial property details and between Rufin VanRullen (VanRullen, 2006) who claimed in a following paper that holistic face processing has only a minor effect on search performance, and that when Fourier amplitude information (which carries global low-level statistical properties of images) is made irrelevant, and only phase information (carrying

contour localization) is used to detect faces, then the results by Hochstein are in fact contradicted. His claim was that the new results imply, contrary to the previous work, that the face pop-out effect is mostly based on low-level factors. That said, both authors agreed about the existence and the high level of pop-out effect by faces. While the debate between Hochstein and VanRullen persisted in two

more consecutive publications debating the reasons for the face attractiveness effect, we in our group got interested mainly in the pure phenomenon of face attraction independent of the task and the distractor.

If faces indeed capture the attention no matter what, and in the same speed, we felt, then this could well be evidence for the fact that saliency and attention allocation for highly semantic objects such as faces could

be affected by a purely bottom-up mechanism based on features of the image. This triggered our interest and made Christof and I meet on a Sunday morning at a small coffee shop on Colorado Boulevard to discuss the option of a possible experiment where, instead of pure reaction-time measurement, we would use eye-tracking as a metric to quantify the attention allocated by observers to images with faces., And, instead of a pop-out effect measured in a grid of competing stimuli we would use real-life natural scenes with faces and try to measure similar effects. Finally, if the effects did repeat themselves in our work, we would argue that we can

THE FACE

An influential British magazine often referred to as the "80s fashion bible", which tried to keep a finger on the pulse of youth culture for over two decades. It was started in May 1980 by Nick Logan out of his publishing house, Wagadon. In the late 1980s it contained the article suggesting that Jason Donovan was gay. By its May 2004 closure the format had become a worldwide standard.

add a module to the saliency map model of Itti and Koch (see Methods chapter) that will use facial information as a measure for attention prediction, among other bottom-up parameters.

A sequence of emails between our group and Ruffin, as well as Laurent Itti from the University of Southern California, led to the start of this study about faces. A study which had various branches in the end and proved to reflect on much more than just the question of “do faces really pop-out, and, if so, can we incorporate this in a saliency map model?” This was the question in the final email sent between Christof and myself. I hope that the coming chapters will shed light on the answer to it.

Prior studies about faces

First, I shall discuss some of the prior studies and known questions/answers in the faces research literature.

Almost all creatures, from dogs to canaries, gorillas to penguins, and certainly members of our human family, manifest the same formula for facial composition: a forehead, two eyes, a nose, a mouth, and a chin that are in the same relative positions. Why was nature so consistent in composing faces? Essentially, the arrangement optimizes survival: the eyes located high for a commanding view of the world, the nose turned down to avoid the rain, and the mouth situated to ingest food that has been perused by the nose and eyes above it.

Throughout the long course of human evolution, recognizing and evaluating people on the basis of their faces made an important contribution to survival. Knowing whom to trust and whom to suspect saved many of our ancestors from needless bloodshed, so they could live to obtain food and mate another day. As cooperative societies emerged and people relied on

each other for working together in hunting, food gathering, or constructing shelter, it was helpful to rely on bodily and facial expressions as much as on words to comprehend intentions and emotions. The formation of alliances, vitally important in what we might call the “hive-life” of social people, depended on communicating trustworthiness, which could be done, in part, through the always-conspicuous face. Ability to read faces became a tool for human survival.

Charles Darwin’s classic book, *The Expression of the Emotions in Man and Animals* (Darwin, 1872), pointed out that facial gestures, as well as some postures, were part of the survival kit of some animals. Discussing aggressive expressions of animals, Darwin wrote: “When a dog is on the point of springing on his antagonist, he utters a savage growl; the ears are pressed closely backwards, and the upper lip is retracted out of the way of his teeth, especially of his canines.”

The early work of Darwin indicated that people all over the world both express and perceive facial expression in similar ways. These early observations were confirmed by Paul Ekman, who enlarged Darwin’s observations, using more sophisticated methods (Ekman, 1973). Traveling to exotic locations such as Papua New Guinea where the residents had little contact with outsiders, Ekman showed the natives photographs of people expressing anger, happiness, disgust, surprise, sadness, and fear and asked them to say what emotion was being expressed. He found wide general agreement in the emotions identified, which he interpreted as meaning “our evolution gives us universal expressions, which tell others some important information about us”. In these universal signals we are able to read another’s emotions, attitudes, and truthfulness.

Current debates regarding faces can be broken into various topics of concern:

Cognitive debates address the pathways by which faces are detected, mainly through the hypothesis for face detection by Bruce and Young (Bruce & Young, 1986) according to which face perception is in fact pathway independent. The key debates argue between faces being detected in an “exemplar” versus a “prototype” method; and between the “holistic” versus “feature-based” method. Supporting the feature-based method are studies about face inversions (where parts of the face are inverted while the eyes and mouth remain in the correct direction, showing that people perceive the person’s identity just as fast as they would in the noninverted case). Studies by Wolfe (Wolfe & Horowitz, 2004), Pessoa (Pessoa, McKenna, Gutierrez, & Ungerleider, 2002), and recently in our group by Reddy and Koch (Reddy, Wilken, & Koch, 2004) argue that faces have special effects on attention based on emotion/gaze-direction or based on demographic cues such as gender or skin color. The studies abovementioned by Ruffin VanRullen versus the ones by Hochstein argue for and against peripheral versus Foveal interaction.

On the neuronal level the debates concern mainly the argue between modular versus distributed face perception. The modular version supported by studies by Nancy Kanwisher and Kalanit Grill-Spector suggest a clear area in the brain, known as the Fusiform Face Area which is the main contributor to face perception. The work of Haxby (Haxby, Hoffman, & Gobbini, 2000) rather suggests a distributed range of areas and mechanisms that provide the complex detection of faces. On the same level, a debate between the group of Kanwisher and the group of Gauthier argue for and against (Gauthier, Skudlarski, Gore, & Anderson, 2000; Kanwisher, 2000) the domain-specific face detection *per se*. The group of Rafi Malach from the Weizmann institute used fMRI studies to argue towards a Foveal perception of faces rather than the earlier suggested Peripheral mechanisms involved in the perception (Levy, Hasson,

Avidan, Hendler, & Malach, 2001). Finally, the role of emotions and expressions in faces are studied by groups such as the Adolphs group at Caltech (R. Adolphs, 1994). Of major interest to my work is a recent thorough study conducted by one of my committee members – Doris Tsao – who actually used electrophysiological stimulation as well as fMRI in monkeys to trace the connectivity between and across multiple patches that seem to be integrative to facial perception. This intriguing work also used a broken-down wide spectrum of features in the face to construct the exact nature of the invariance to a particular person versus the invariance to a particular facial orientation or facial relation between parts of the face (nose, eyebrows, etc.) (Tsao, Freiwald, Tootell, & Livingstone, 2006; Tsao, Moeller, & Freiwald, 2008).

On the computational level various mechanisms are provided to support rapid face detection and face recognition, sometimes but not necessarily inspired by biological mechanisms in the brain. Of profound research are the famous algorithm by Viola and Jones which will be discussed in length in a later paragraph, but also algorithms such as SIFT, or other feature-detector-based algorithms which provide remarkable results in face detection. Suggested algorithms based on face-space decomposition suggest ways to rapidly identify faces using 2D images containing face features that are easy to separate faces from noise by, such as the eigenface. 2D cartoon images are used to devise extreme faces and anti-faces profiles, or to identify face profiles. Finally, recent works focus on developing algorithms for face classifications, such as gender or race, using support vector machines. Hybrid models of the abovementioned using algorithms such as the Hmax family or the saliency model assist in identifying the role of low-level versus high-level features used for face detection.

To sum these algorithms and methods, following is a table with the relevant processes and analysis methods used for face detection and identification:

	Computational	Cognitive	Neural
Detection	Rapid Object Detection Using a Boosted Cascade of Simple Features (Viola & Jones, 2001)	Looking upside-down faces (Yim, 1969)	Representation of Visual Stimuli in Inferior Temporal Cortex (Gross & Schonlen, 1992)
Representation	Low-dimensional procedure for the characterization of human faces (Sirovich & Kirby, 1987)	Face-space models of face recognition (Valentine & Cross, 2001)	Prototype-referenced shape encoding revealed by high-level aftereffects (Leopold, O'Toole, Vetter, & Blanz, 2001)
Categorization	Detecting faces in images: A survey. (Yang, Kriegman, & Ahuja, 2002)	Adaptation to natural facial categories (Webster, Kaping, Mizokami, & Duhamel, 2004)	Shape analysis of female facial attractiveness (Valenzano, Memucci, Tartarelli, & Cellerino, 2006)
Identification	Face recognition using eigenfaces (Turk & Pentland, 1991)	Understanding face recognition (Bruce & Young, 1986)	Invariant visual representation by single neurons in the human brain (Quiroga, Reddy, Kreiman, Koch, & Fried, 2005)
Social context			Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala (R. Adolphs, 1994)

Table 1 - Summary of methods and processes for faces detection

Finally, based on a review of “19 results all computer vision research should know about faces” (Sinha, Balas, Ostrovsky, & Russell, 2006) here are the major results relevant to my following research out of these:

1. Humans can recognize faces in extremely low-resolution images

Unlike current machine-based systems, human observers are able to handle significant degradations in face images. For instance, subjects are able to recognize more than half of all familiar faces shown to them at the resolution depicted in Figure 13.



Figure 13 - Subjects are able to recognize familiar faces in low resolution

The individuals shown from left to right are: Prince Charles, Woody Allen, Bill Clinton, Saddam Hussein, Richard Nixon, and Princess Diana. From (Yip & Sinha, 2002).

This result is important for us as we used varying contrast, various face sizes, and various focus types and lenses in our image dataset. Although our images tend to show the faces clearly, we use this result to claim that even in the more blurry images the existence of a face and the ability for our subjects to identify it are evident.

2. Humans can recognize faces in extremely low-resolution images

Images which contain exclusively contour information are very difficult to recognize, suggesting that high-spatial frequency information, by itself, is not an adequate cue for human face recognition processes.

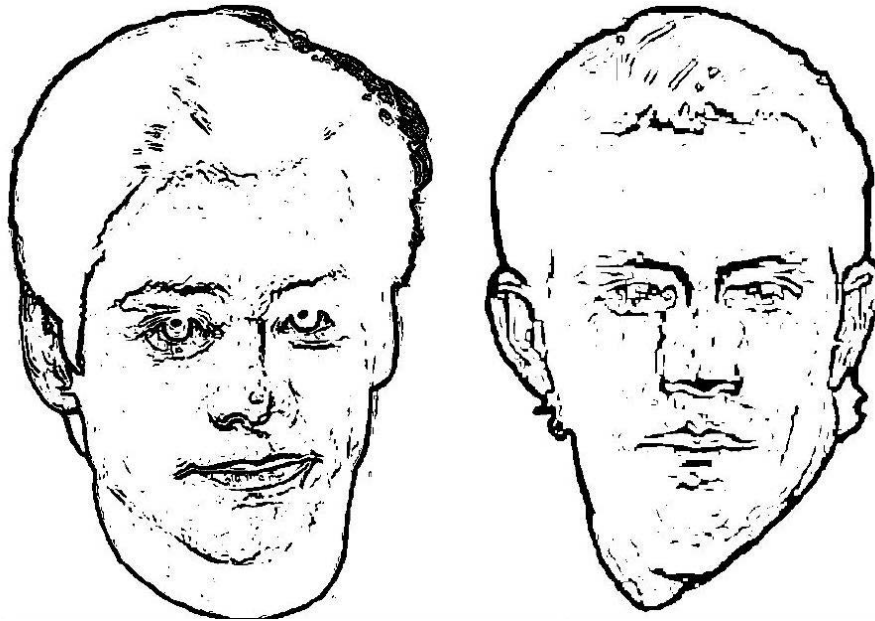


Figure 14 - Images with only contour information difficult to recognize

Shown here are Jim Carrey (left) and Kevin Costner. From (Pearson, Hanna, & Martinez, 1990)

While none of our images were that low resolution, we did use some black and white images, and were considering even using cartoon faces – especially for the Prosopagnosia experiments – and therefore wanted to make sure that faces are indeed recognized as such even in the extreme cases where they are stripped of all the features that usually allow for recognition.

3. Facial features are processed holistically

Subjects find it more difficult to name the famous faces depicted in the two halves of the aligned halves image than the misaligned because holistic processing interacts (and in this case, interferes) with feature-based processing.



Figure 15 - Holistic versus feature-based processing

Naming Elvis from John F. Kennedy is easier when the two half faces are not aligned — breaking the holistic perception of the face into its parts. From (Calder, Young, Keane, & Dean, 2000)

As we argue in our work towards a bottom-up basis for the detection and allocation of attention towards faces this result was important for our understanding of the literature concerning face perception and recognition, as we were interested in understanding whether and if so which parts of the faces are the ones governing the perception of these as such. We used this information to construct an image set with faces where features such as nose/eyes/mouth were scrambled to demonstrate that subjects still perceive these as faces and continue to attend to them rapidly.

4. Both internal and external facial cues are important and they exhibit nonlinear interactions

While it is typically assumed that internal features (eyes, nose, and mouth) in the face and their mutual spatial configuration, are the critical constituents of a face, and the external features (hair and jaw-line) are too variable to be practically useful, it is in fact not necessarily the case. In a study where the internal features were the same, but the external ones differed, subjects could rapidly recognize identities of people based purely on the external differences.



Figure 16 - Identity recognition based on purely external cues

Although this image appears to be a fairly run-of-the-mill picture of Bill Clinton and Al Gore, a closer inspection reveals that both men have been digitally given identical inner face features and their mutual configuration. Only the external features are different. It appears, therefore, that the human visual system makes strong use of the overall head shape in order to determine facial identity. From (Sinha & Poggio, 1996)

In a similar fashion to that pointed out in the previous section bullet we used this information regarding the perception of faces based on external cues to alter some of the images and construct a dataset for testing of viewing of abnormal faces.

5. The configural relationships is independent across width and height

Even drastic compressions of faces do not render them unrecognizable. Celebrity faces which have been compressed to 25% of their original width did not affect recognition performance in the same set as was obtained with the original faces.



Figure 17 - Changes in configural relationship in faces do not hurt perception.

In the following images, the compression across one axis does not hurt the ability to recognize the identities. From left: Ronald Reagan, Jason Alexander, Prince Charles, George W. Bush, Robin Williams, and Woodie Allen

In a similar fashion to that pointed out in the previous bullets we used this information regarding the perception of faces based being independent on ration across features and resolution to alter some of the images and construct a dataset for testing of viewing of abnormal faces. Mostly this was used for us as a control for the claim that none of the image features *per se* is the key attractor to faces.

6. Vertical inversion dramatically reduces recognition performance

Images in which the eyes and mouth have been vertically inverted, show higher decrease in recognition performance than ones where all the features but the eyes and mouth were inverted.

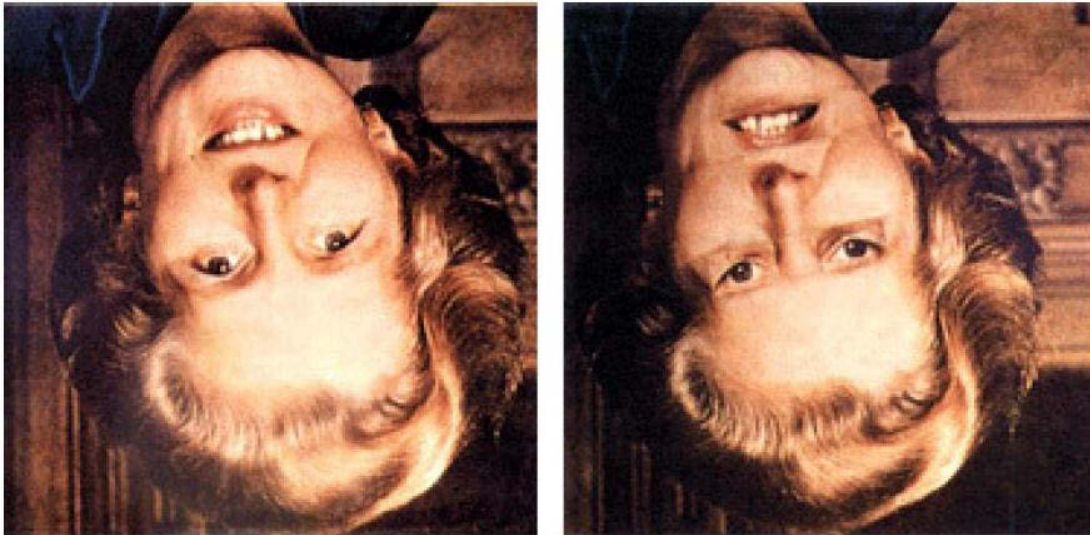


Figure 18 - Inverted faces: The Thatcher illusion

When the whole face is inverted (**right**), this manipulation is not apparent. If the reader turns this page around, however, the manipulation is grotesquely obvious. From (P. Thompson, Facial, Famous, & Optical, 1980)

7. Contrast polarity inversion dramatically impairs recognition

Faces are particularly difficult to recognize when viewed in reversed contrast, as in photographic negatives. Though no information is lost, our ability to use the information in the image is severely compromised. This suggests that some normally useful information is rendered unusable by negation.



Figure 19 - Negative contrast image from a Beatles' album

This image contains numerous well-known celebrities, whose likenesses would be easily recognizable to many readers of this publication. However, when presented in negative contrast, it is difficult, if not impossible, to recognize most of the faces. From (Bruce & Langton, 1994)

8. The visual system starts with a rudimentary preference for face-like patterns

Newborns selectively gaze at “face-like” patterns only hours after birth.



Figure 20 - Newborns preferentially orient their gaze to the face-like patterns.

Top image captures newborns' attention more than **bottom** one, suggesting some innately specified representation for faces. From (Johnson, Dziurawiec, Ellis, & Morton, 1991)

Since we later argue towards an innate biological brain mechanism that governs the bottom-up attention allocation to faces, we based our hypothesis on the claim that even newborns already show the desire to attend to faces as a suggestive evidence to these being indeed an internal brain mechanisms that we are born with.

9. The human visual system appears to devote specialized neural resources for face perception

Primary locus for human face processing may be found on the fusiform gyrus of the extra-striate visual cortex (Kanwisher, McDermott, & Chun, 1997). This region shows a pattern of selectivity suggesting a strong domain-specific response for faces. The characterization of the “fusiform face area” (FFA) as a dedicated face processing module appears very strong.

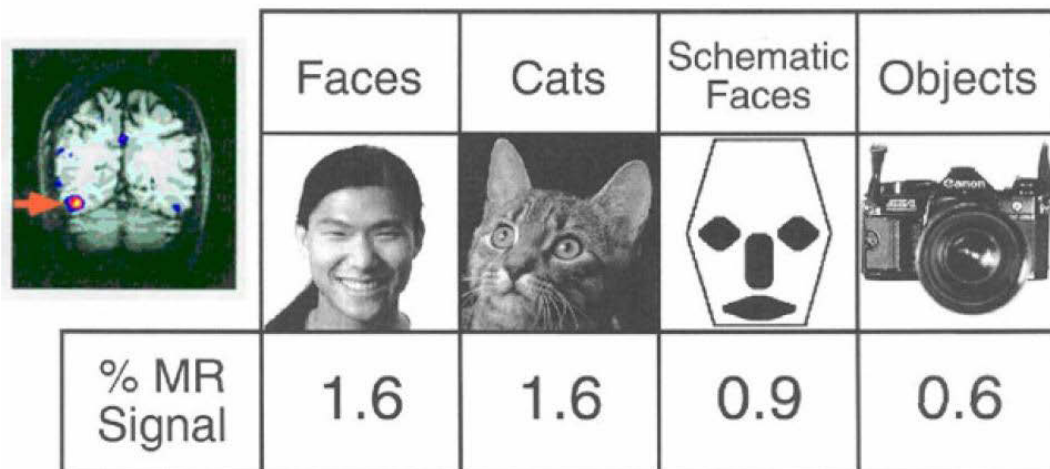


Figure 21 - An example of the FFA in one subject

Showing right-hemisphere lateralization (Tong, 2000) together with the amount of percent signal change observed in the FFA for each type of image. Photographs of human and animal faces elicit strong responses, while schematic faces and objects do not.

Since our work suggest that faces are indeed attended to rapidly and mostly due to an unbiased attention allocation mechanism that is bottom-up based, the existence of a brain region governing face perception can be used as a strong evidence to support the fact that the brain indeed treats faces differently than other highly semantic learned cues.

10. Latency of responses to faces in IT (120 ms) suggest a feed-forward computation

There is neurophysiological evidence that truly complex tasks, such as face recognition, may be carried out over a surprisingly short period of time. The computational relevance of these results is that recognition as it is performed up to the level of IT cortex probably requires only one feed-forward pass through the visual system. Feedback and iterative processing are likely not major factors in the responses recorded in these studies. This is a very important constraint on recognition algorithms, as it indicates that sufficient information must be extracted immediately from the image without the luxury of resorting to slowly converging iterative computations.

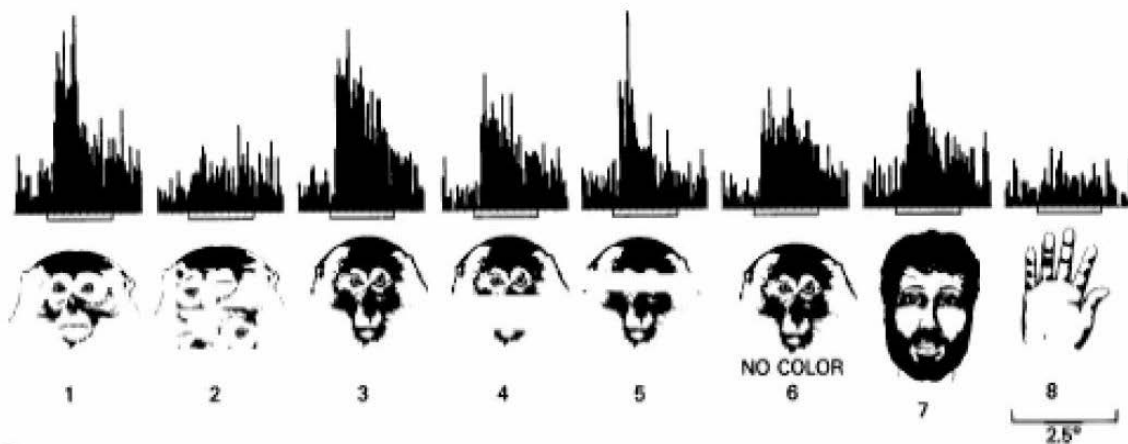


Figure 22 - Examples of an IT cell's responses to variations on a face stimulus

The response is robust to many degradations of the primate face and also responds very well to a human face. The lack of a response to the hand indicates that this cell is not just interested in body parts, but is specific to faces. Cells in IT cortex can produce responses such as these with a latency of about 120 ms. From (Desimone, Albright, Gross, & Bruce, 1984).

The rapid neuronal activity towards looking at faces serves as a baseline for our timing testing of the amount of time it takes subjects to either view a face when freely viewing a scene, to view a face when looking for one, or to view a face when trying not to.

Face detection

Lately, face detection has become such a profound and important feature of our daily usage. Most conventional point and shoot cameras nowadays includes a built-in module that provides immediate detection of faces in the images made to assist the photographer in centering and correctly aligning the camera before taking a picture. Most search engines now include a module that allows the searcher to directly seek images that include faces in them, implying that a face detection mechanism is typically built into every search.

While there are various algorithms and methods for computer-based face detection, my work mainly put to use a very common and standard one that I will discuss shortly in the following paragraph: The Viola and Jones face detection algorithm (Viola & Jones, 2001).

The Viola-Jones object detection framework is the first object detection framework to provide competitive object detection rates in real-time. One of its main benefits is its rapid processing,

which allows it to be used commonly for real-time face detection by web-cameras, for example. In order to reach this speedy detection the Viola and Jones utilizes 3 mechanisms: images are represented efficiently using a mechanism where each pixel acts as an intermediate representation of the preceding pixels in a rectangle marked by the x,y coordinate of the image. Secondly, an efficient learning method creates a large number of quantifiers for future detection. Third, a smart

Face

T. Templeton "Faceman" (usually referred to simply as "Face") is the smooth-talking con-man who serves as the team's appropriator of vehicles and other useful items in the legendary 1980s TV series *THE A-TEAM*.



cascading window technique allows for single-run detection.

Once the image was aggregated using the summed rectangles “features” are extracted based on the training data. The features employed by the detection framework universally involve the sums of image pixels within rectangular areas. As such, they bear some resemblance to Haar basis functions, which have been used previously in the realm of image-based object detection.

However, since the features used by Viola and Jones all rely on more than one rectangular area, they are generally more complex. Figure 23 illustrates the types of features used in the framework. The value of any given feature is always simply the sum of the pixels within clear rectangles subtracted from the sum of the pixels within shaded rectangles.

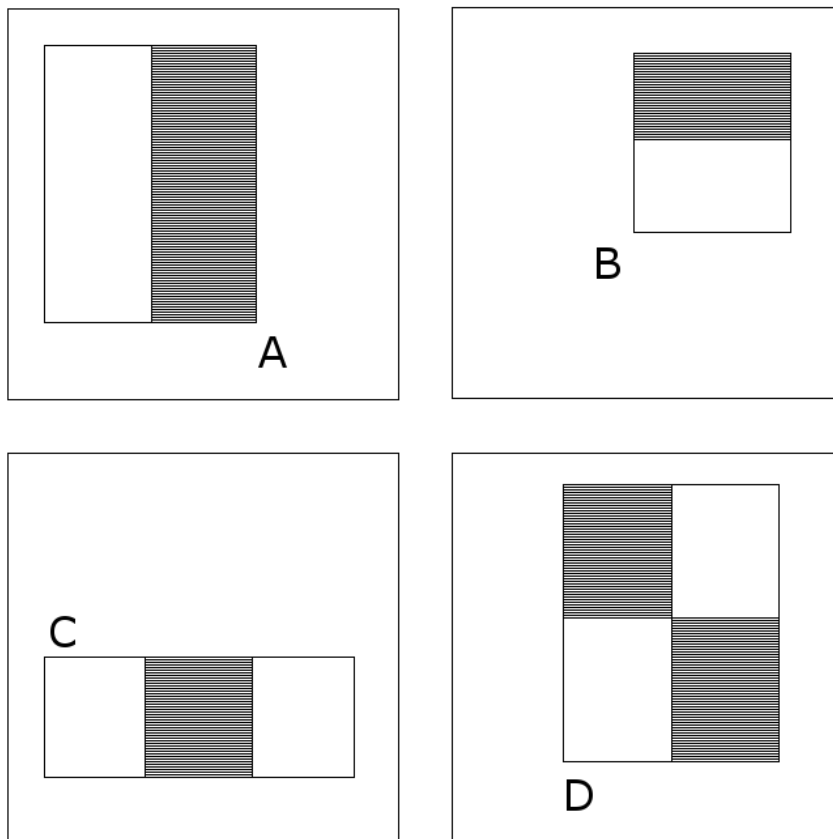


Figure 23 - Feature types used by the Viola and Jones algorithm

Because each rectangular area in a feature is always adjacent to at least one other rectangle, it follows that any two-rectangle feature can be computed in six array references, any three-rectangle feature in eight, and any four-rectangle feature in just nine.

The speed with which features may be evaluated does not adequately compensate for their number, however. For example, in a standard 24x24 pixel sub-window, there are a total of 45,396 possible features, and it would be prohibitively expensive to evaluate them all. Thus, the object detection framework employs a variant of the learning algorithm AdaBoost (Freund, Schapire, & Abe, 1999) to both select the best features and to train classifiers that use them.

The AdaBoost basically combines weak learners in successive iterations with the unlearned example data given higher weights. Using an exponential decrease of training error, it gives good generalization with high margins which allows for quickly discarding of large number of bad features.

While features selected do not necessarily correspond to facial features, interpretations of them allow for considering eyebrows, nose/eyes contrast as significant features in the detection (see Figure 24).

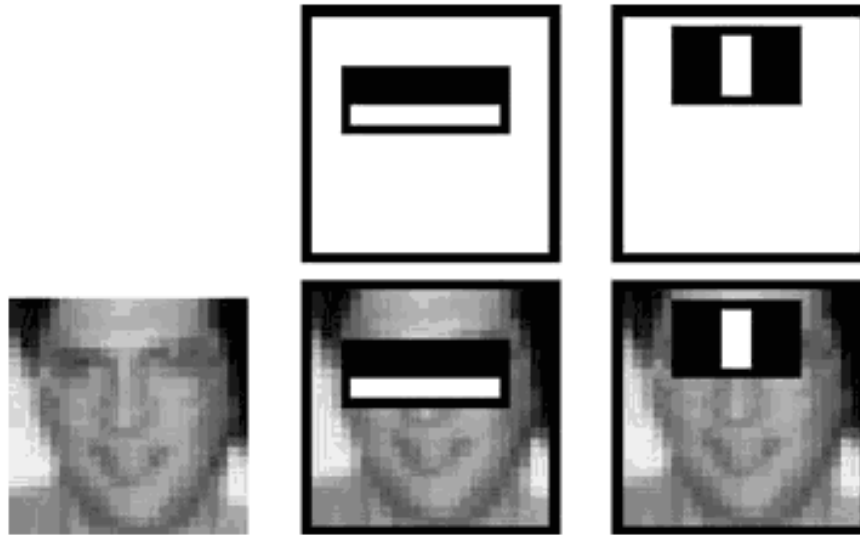


Figure 24 - Features selected by the AdaBoost

The two features are shown in the top row and then overlaid on a typical training face in the bottom row. The first feature measures the difference in intensity between the region of the eyes and a region across the upper cheeks. The feature capitalizes on the observation that the eye region is often darker than the cheeks. The second feature compares the intensities in the eye regions to the intensity across the bridge of the nose.

The evaluation of the strong classifiers generated by the learning process can be done quickly, but it isn't fast enough to run in real-time. For this reason, the strong classifiers are arranged in a cascade in order of complexity, where each successive classifier is trained only on those examples which pass through the preceding classifiers. If at any point in the cascade a classifier rejects the sub-window under inspection, no further processing is performed and the search moves on to the next sub-window (see Figure 25). The cascade therefore has the form of a degenerate decision tree. In the case of faces, the first classifier in the cascade – called the attentional operator – uses only two features to achieve a false negative rate of approximately

0% and a false positive rate of 40%. The effect of this single classifier is to reduce by roughly half the number of times the entire cascade is evaluated.

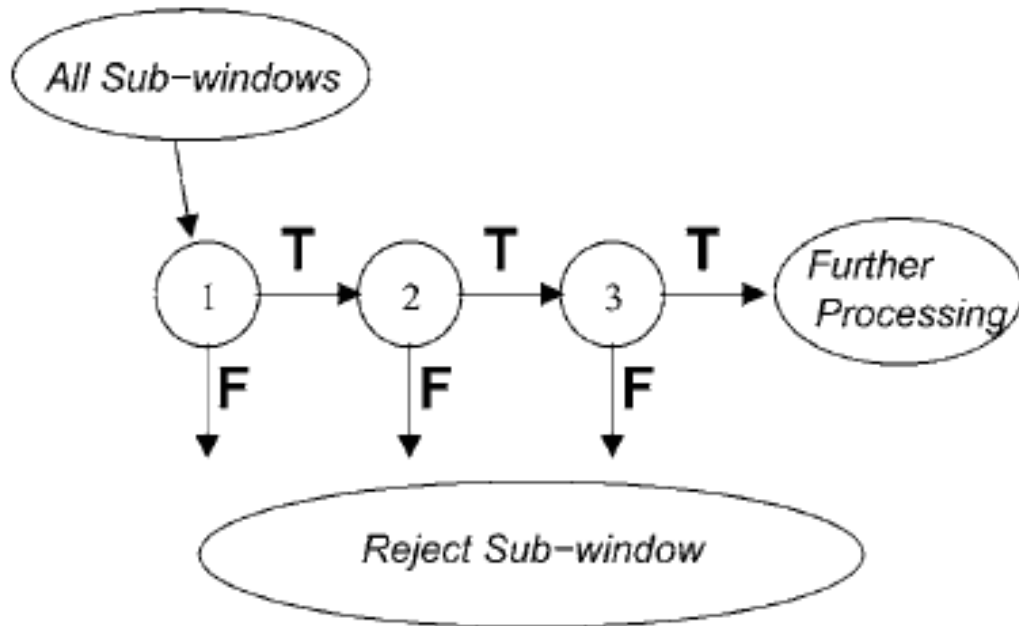


Figure 25 - Cascade architecture of the Viola and Jones algorithm

Multi-pass processing of image allows for high detection, with high false positive rates, but the repeated evaluation of sub-windows allows for a decrease in the number of false positives.

To match the false positive rates typically achieved by other detectors, each classifier can get away with having surprisingly poor performance. For example, for a 32-stage cascade to achieve a false positive rate of 10^{-6} , each classifier need only achieve a false positive rate of about 65%. At the same time, however, each classifier needs to be exceptionally capable if it is to achieve adequate detection rates. For example, to achieve a detection rate of about 90%, each classifier in the aforementioned cascade needs to achieve a detection rate of approximately 99.7%.

In the examples given in the paper describing the algorithm both its speed and its accuracy were measured showing low false positives (0% miss rate, 80% correct rejection rate) and very high speed (near real-time). This is under the limitation of usage of only frontal view training set, with up to 15° in plane, 45° out of plane.

Results from running the Viola and Jones algorithm on the standard set available with the *OpenCV* application are shown below (Figure 26).

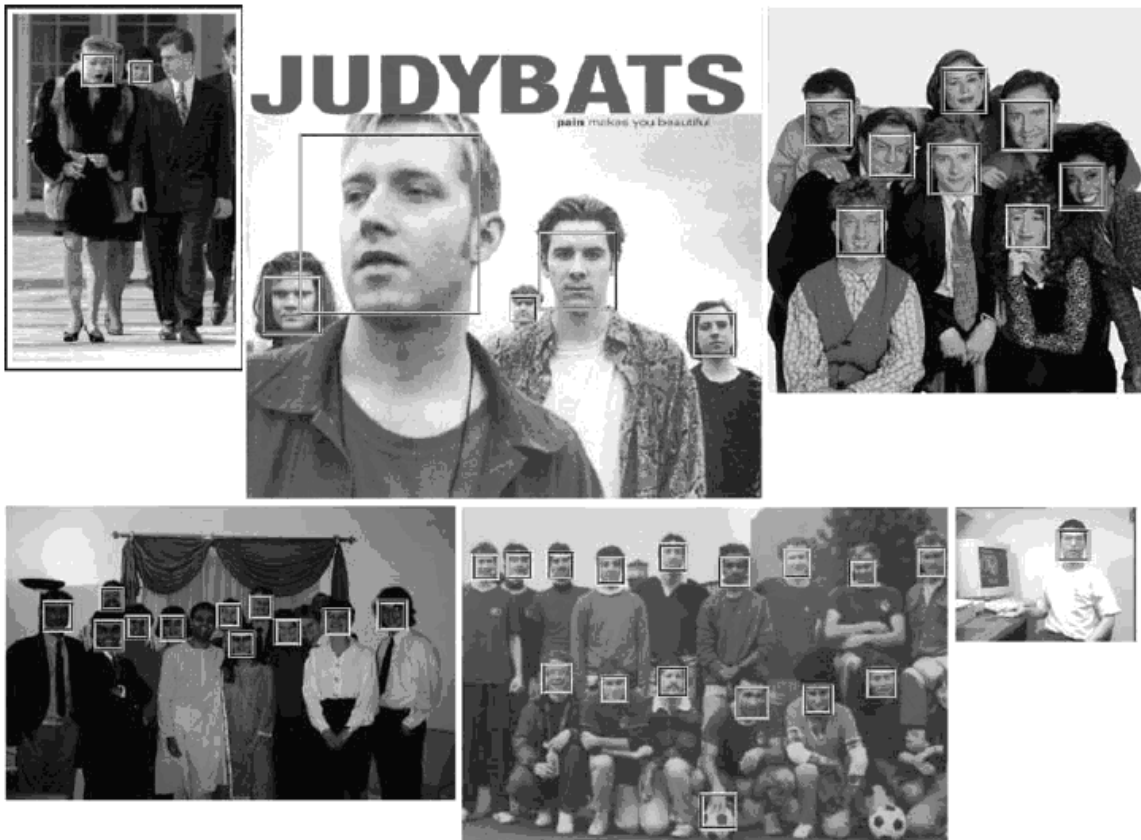


Figure 26 - Examples of the results of the Viola and Jones algorithm

More complex images from our dataset show some more interesting false positives, as well as interesting test where we tried to see what the algorithm would show in complex images. Notice the “devil’s face” detected correctly by the algorithm in high resolution (Figure 27).



Figure 27 - Examples of the Viola and Jones results on complex images

Top left. Image from our dataset used in the following experiments. Notice that for each image we output the coordinates of the center of the face, radius or diagonal of circle/rectangle around the face, confidence level (based on the amount of steps through which the algorithm passed before reaching a decision), and the order by which the faces were detected.

Top right. Image of an ape holding a face-like bag. The ape's face was detected while the bag's "face" was not.

Lower left. The CNN picture of the "devil in the smoke". The algorithm detected the face (as did very many Christian evangelists).

Lower right. A complex illusion with two combined faces which we ran through the algorithm to see what would come out. Interestingly, the algorithm found only a single face among the two options.



Methods

*You can't do today's job with
yesterday's methods and be in
business tomorrow.*

Warren Buffet

During my research I employed various methods and techniques both for the analysis and the experimental research. This chapter will list these methods and techniques in a single area and thereafter will be referred to whenever one of the following methods is mentioned and employed.

Each chapter will just discuss the deviations or additions to the common techniques mentioned here, which include: eye-tracking data, single-neuron recording, local field potential analysis, and computation modeling.

Psychophysics metrics

All data pertaining to reaction times, answers analysis, or data pertaining to pure psychophysics, with no eye-tracking or brain recordings were done either with the Matlab psychophysics toolbox (Brainard, 1996), or – for the study in chapter 9 – using TCL/TK and C. The abovementioned toolboxes record data and timing answer to the level and accuracy of milliseconds and use the internal computer clock, or the screen refresh rate to synchronize the timing. All analysis using the toolboxes were done using the *KbCheck* and *WaitSecs* functions that give higher accuracy for the timing.

Eye tracking

Device

Eye tracking is a method to tap into what people attend to. Although eye tracking was conducted even in the late 19 century using crude painful methods, it mainly dates back as an experimental procedure to the 1950s first eye-tracking experiments conducted by the Russian scientist Alfred Yarbus. Yarbus had subjects look at various scenes and perform different task such as “judge who the visitor is”, or “what do people attend to in the scene” and saw differences in the viewing patters due to the tasks (see Figure 28).

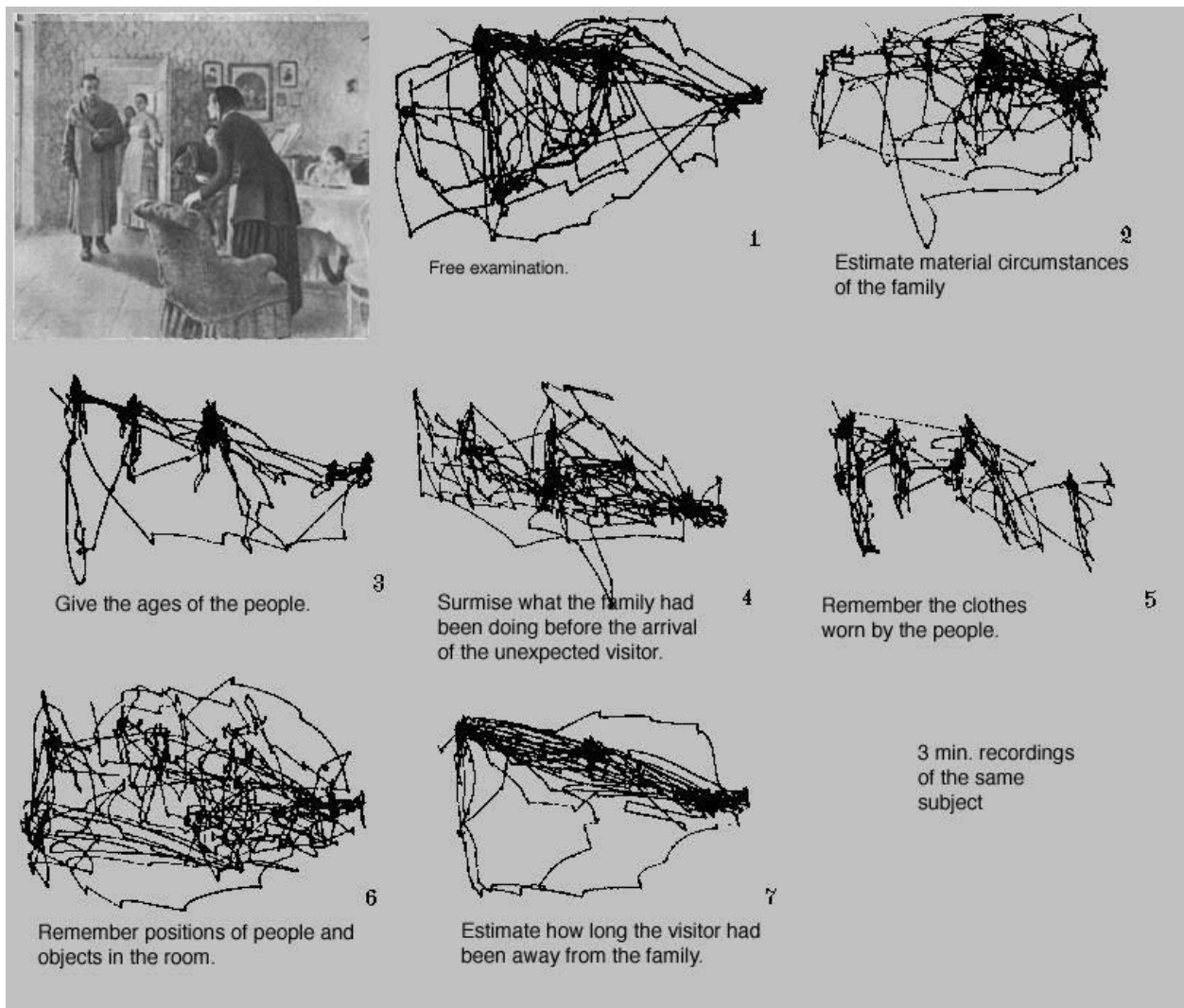


Figure 28 - Example of different eye saccades due to different task

Image taken from (Yarbus, 1967)

The most widely used current eye-trackers are video-based. A camera focuses on one or both eyes and records their movement as the viewer looks at some kind of stimulus on a computer screen. Most modern eye-trackers use contrast to locate the center of the pupil and use infrared and near-infrared noncollimated light to create a corneal reflection. The vector

between these two features can be used to compute gaze intersection with a surface after a simple calibration for an individual.

I used the EyeLink-1000 eye tracker for all my experiments. The EyeLink-1000 tower-based tracker (see Figure 29) sits 80 cm from a CRT2 screen, in a dark quiet room.

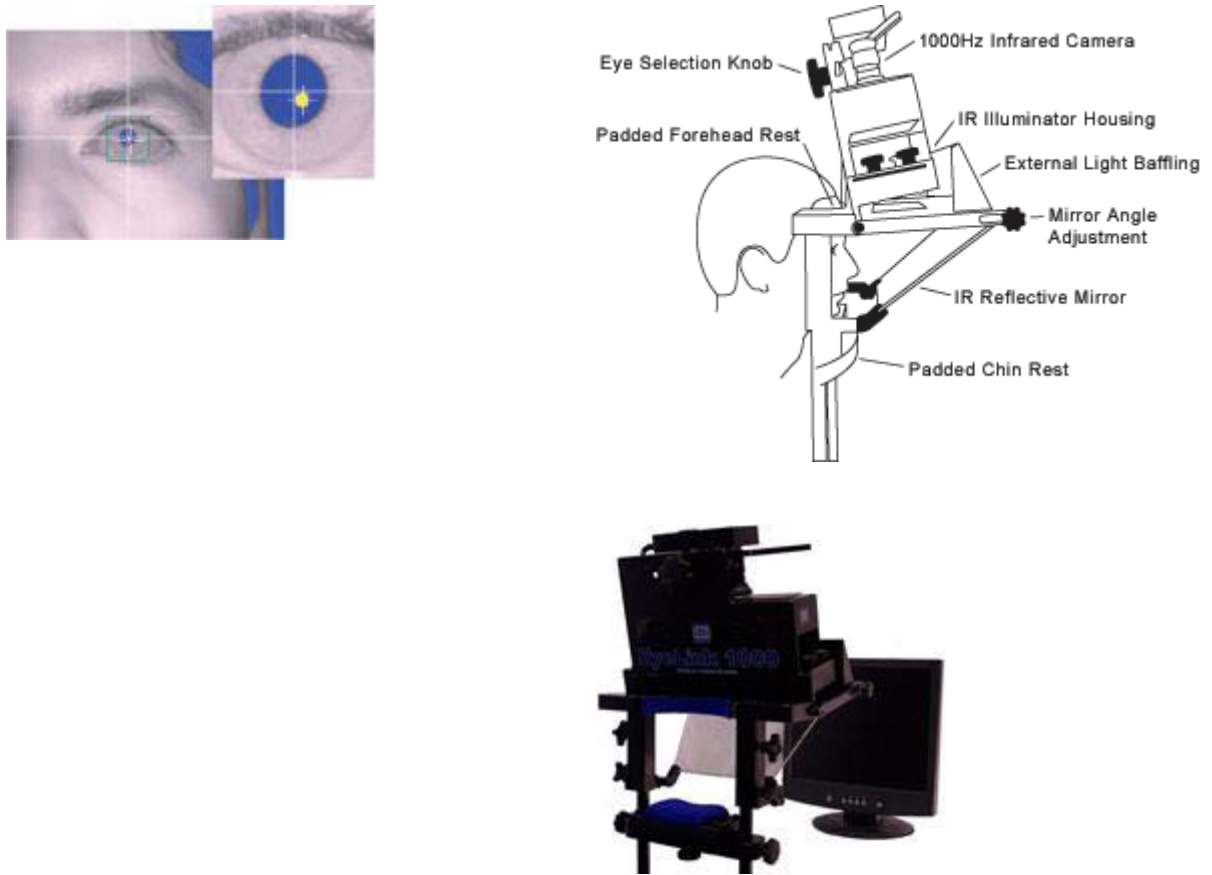


Figure 29 - Eye tracker

Top right. Diagram of the Eye-link 1000 system

Top left. A screenshot of the eye-tracker pupil and eye images

Bottom. Pictures of the eye-tracking system

I used a chinrest to support the head, which allows for lower drift during the experiment. The eye-tracker samples the eyes at 1000 Hz and collects the x,y coordinates of the eyes at each timepoint. On top of acquisition the system performs fixations detection.

Each tracking session was initiated with a calibration and validation session where a participant performed a repeated target fixation task to 15 screen locations, in random order (see Figure 30).

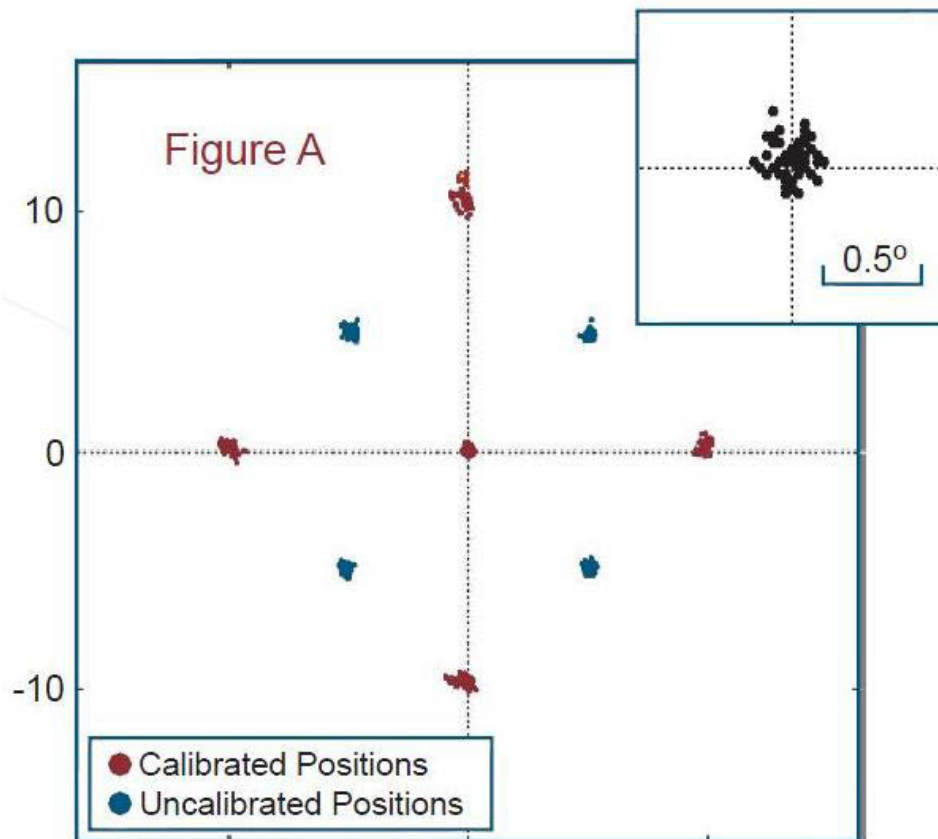


Figure 30 - Fixations accuracy in the eye-tracker

A diagram of the fixations accuracy using the EYELINK-1000. Subjects first fixated in each of the 9 locations, and later were asked to fixate on the same location. Shown here are the clusters of

fixations by various subjects on the same 9 locations, with the average error of 0.5° demonstrated for the center fixation in the small box.

Following is a table with the specifications of the Eyelink-1000 parameters used in my experiments:

Monocular sampling rate	1000 Hz
Eye tracking principle	Pupil with corneal reflection
Average accuracy	0.25° to 0.5° typical
Saccade event resolution	0.05° microsaccades
Spatial resolution	0.01° @ 1000 Hz
Sample delay	Up to 1.8 ms
Blink recovery time	1 ms
Pupil detection model	Ellipse fitting
Gaze tracking range	60° horizontally, 40° vertically
Allowable head movement	25 x 25 x 10 mm (horizontal x vertical x depth)
Camera-eye distance	38 cm
Glasses compatibility	Good
Infrared wavelength	910 nm

Table 2 - Eyelink-1000 parameters

Images

All the images for the experiments conducted were part of a 940 image dataset created specifically for the purpose of the eye-tracking studies. The images were all 1024 x 768 pixels wide. Images were taken such that color spectrum would not saturate, to allow for better

comparison to the saliency model. Images were taken indoors and outdoors and included mainly people and faces for the study of subjects' attention to faces. Faces images mainly contained pictures of myself and my neighbors and friends, but included various other pictures of lab members and volunteers. This allowed us to encapsulate a large variety of ages (from babies to elders), skin colors (black, white, asian, hispanic), emotions (smiling, saddened, disgusted, angry, surprised, neutral), body types (full body, full face), etc. All the images are available online as a dataset on <http://www.klab.caltech.edu/~moran/db/faces>.

The images were introduced as “regular images that one can expect to find in an everyday personal photo album”. Scenes were indoors and outdoors still images (see examples in Figure 46). Images included faces in various skin colors, age groups, and positions (no image had the face at the center as this was the starting fixation location in all trials). A few images had face-like objects (see balloon Figure 46, panel 3), animal faces, and objects that had irregular faces in them (masks, the Egyptian sphinx face, etc.). Faces also varied in size (percentage of the entire image). The average face was 5%-1% (mean \pm s.t.d.) of the entire image. (between 1° to 5° of the visual field); we also varied the number of faces in the image between 1-6, with a mean of 1.1 ± 0.48 . Image order was randomized throughout all experiments. Subjects fixated on a cross in the center before each image onset.

The experimental paradigms making the eye-tracking studies were made of 4 blocks. Block 1 was a “free viewing” of a set of 200 images from the pool (always the same subset).

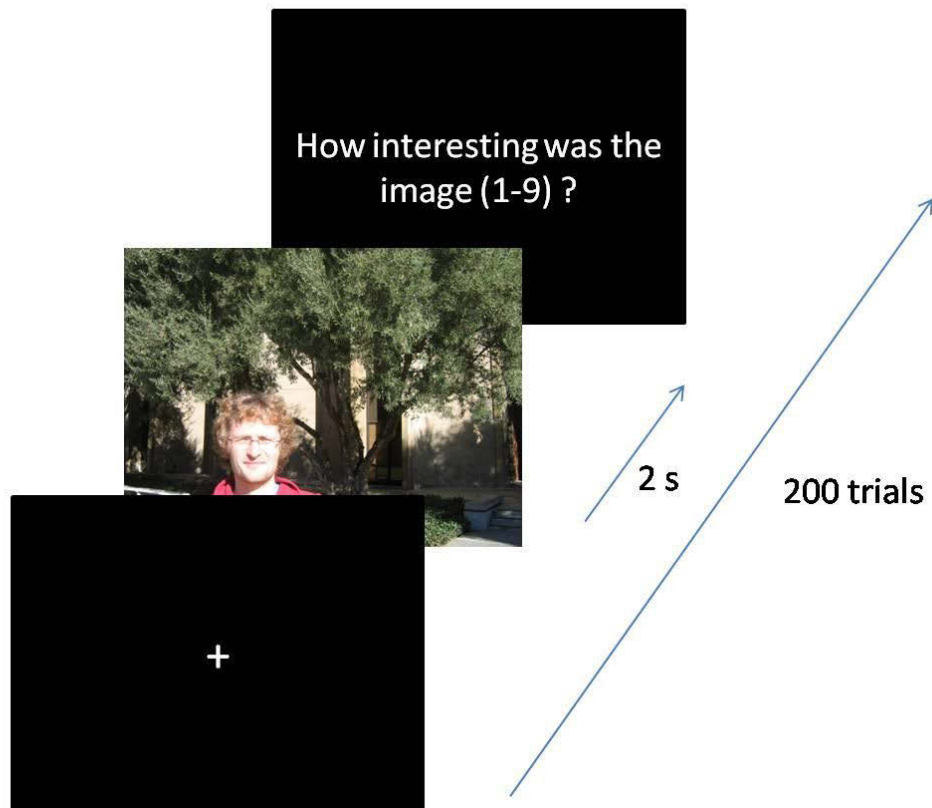


Figure 31 - Illustration of the vision experiment: block 1. Free viewing

Block 2 included 200 images (some of which are repeated and some new), while the task was a “search task”. Subjects were seeing a probe - one of nine objects: cell phone, phone, Rubik’s cube, toy fire-truck or a banana or people’s faces (see Acknowledgment for an elaborate discussion of the people: Yadin, Or, Moran, or Yoni) for 600 ms and afterwards saw the image, in which they had to tell if the probe was in the image or not.

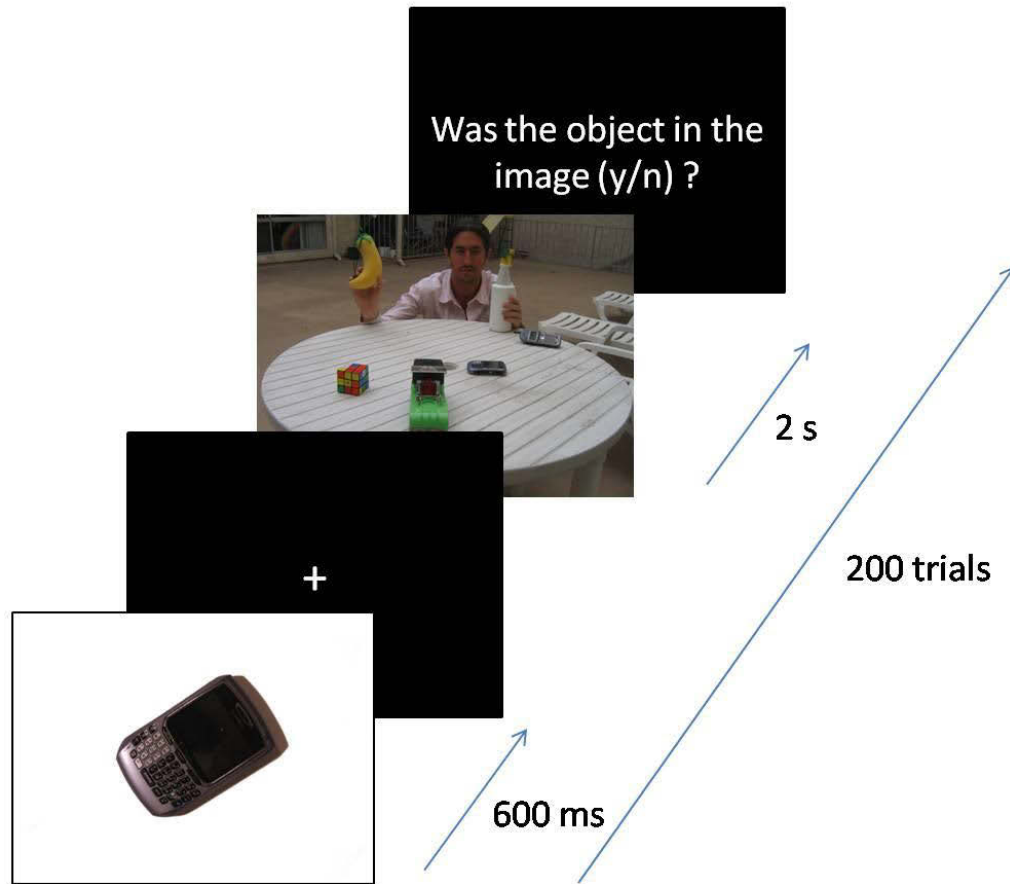


Figure 32 - Illustration of the vision experiment: block 2. Search task

Block 3 was a repeated free-viewing task with 200 images, some new, some old (as a measure of changes in the viewing due to repetition), and some old/new in black and white (in order to test for the effects of color on saliency and attention/fixations predictions).

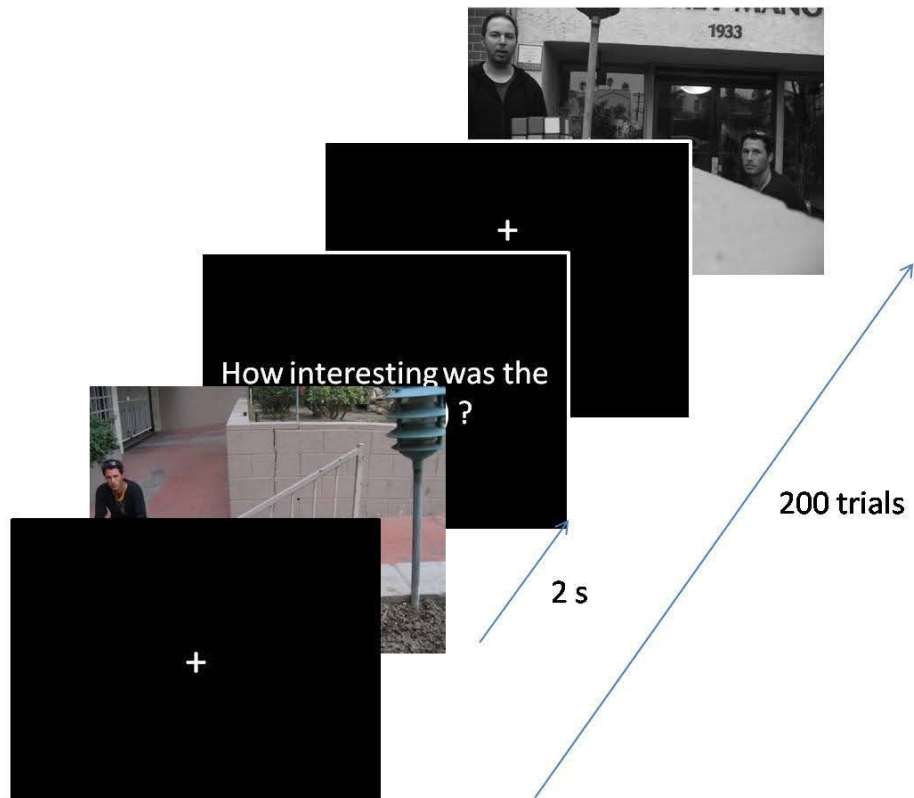


Figure 33 - Illustration of the vision experiment: block 3. Free viewing

Block 4 included 100 images – 50 new and 50 old, from the previous 3 blocks – where subjects were to tell if they had seen the image before. This block was a recognition memory block in order to determine if subjects were attentive during the previous tasks. If they were not, we would expect lower performance in being able to recognize previously viewed images.

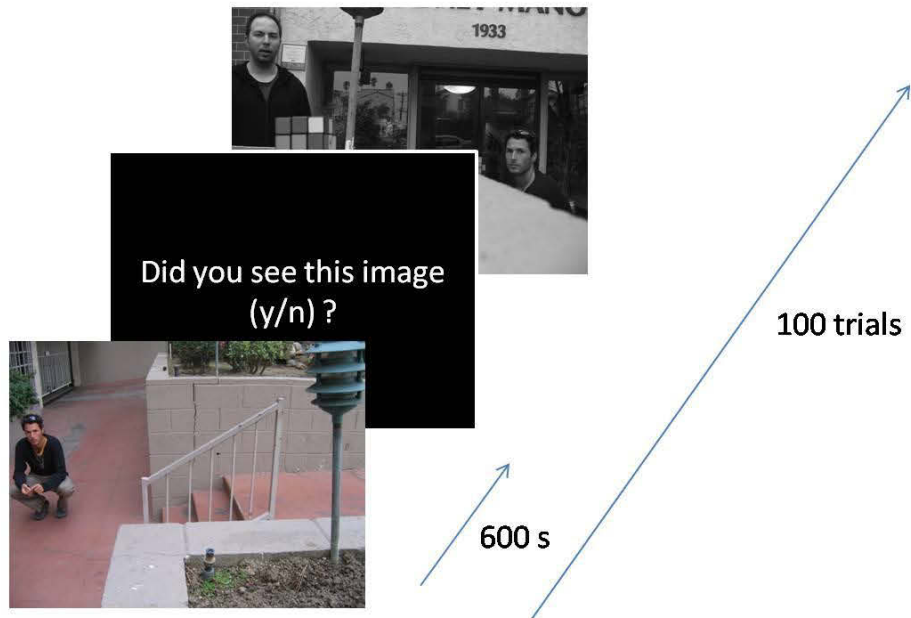


Figure 34 - Illustration of the vision experiment: block 4, Memory block

Figure 35 shows an illustration of the images distributions across tasks.

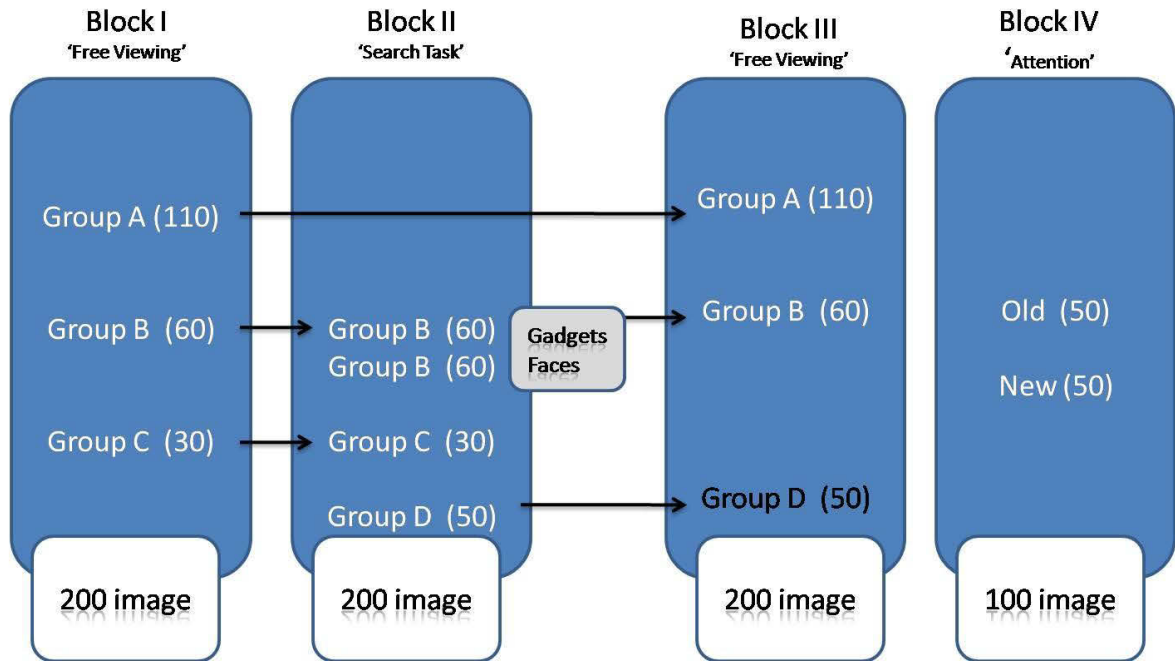


Figure 35 - Images distribution across tasks

Black arrows mark repetition of the same image block in the following block.

Presentation times

Images were presented in various timings, relative to the screen refresh rate of 120 Hz. Most often images were presented for 2 seconds as this was shown in one of the study to yield high enough accuracy and consistency in the answers, be long enough for people to pay attention to details and remember the content, as well as allow us to collect approximately 7 fixations from each image. Finally, though, this was short enough to allow us to have subjects view a sequence of approximately 200 images in a block, taking about 15 minutes for each block, thus not boring or tiring the subjects.

Exposure

As the images in the experiment were later used for comparison with the saliency model we wanted to control for the color spectrum in the images and made sure none of them were either over- or underexposed. Over-exposed regions in images, for instance, will generate higher results in the intensity channel of the saliency model and will end up with stronger weights in the saliency map not necessarily due to the actual content.

Most of the common point-and-shoot cameras that are mainly used to take pictures of people and make sure their faces are nicely lit in the image introduce exposures deliberately.

Indecent exposure is the deliberate exposure by a person of a portion or portions of his or her own body under circumstances where such an exposure is likely to be deemed an offense against prevalent standards of decency and may in fact be a violation of law.

US Federal Law

Taking a random set of images from a commonly used digital camera like the ones we would typically use for normal picture taking would reveal that many of the images taken include some regions that are either overexposed or underexposed. Professional photographers taking images of people and wanting to emphasize faces in them would deliberately overexpose parts of the face so features in it would be nicely lit. Generally, we are used to seeing images that are either overexposed or underexposed daily, but our brain easily compensates for the shifts in exposure and remains invariant to those minor changes. This is true to the extent that upon asking a common observer to take an image that is completely normalized (not underexposed and not overexposed at any region) would be a much harder task than one thinks. Professional photographers are trained in controlling the camera exposure in order to make sure the image is indeed normalized, and — in fact — there are very many scenarios where taking such a picture is impossible. If the frame contains very high contrast between adjacent regions (a dark-

skinned man standing and the sun reflecting strongly just behind him for instance), it is nearly impossible to have an image that would have utterly neutral exposure.

Having said that, we can imagine that many computer vision models which are based on colors, luminosity, and contrast in the image would be severely impaired and vulnerable to exposure effects.

In order to control for the exposure we used following method:

The majority of the images used in the experiment were taken with a digital SLR camera (Konica Minolta Dynax 5D, 6 Megapixels, 1 CCD, Exposure Compensation: -2.0 to +2.0 EV in 0.3 EV steps⁵). The images were taken using a tripod with little to minimal effects of camera movement (when possible, images were taken using a 2 second delay from the time of button-click to eliminate any camera movement due to the button-press itself; in all other cases, as little movement as possible was used). After each image was taken, the color histogram on the camera was presented and the scenes were set up such that the frames would have little to no contrast changes so that the images would have no over- or underexposed areas. The color histograms were tested after each image taking to make sure that no area in the frame was either over- or underexposed.

For all the images we separated each image to its 3 color components (red, green, blue) and made sure that the histograms covered the entire dynamic spectrum (e.g., if the brightest color in the red map wasn't 255 but 217 we multiplied the values by 255/217 so that the map

⁵ Details: <http://www.dpreview.com/news/0507/05071503kmmmaxum5d.asp>

covered the entire spectrum). We multiplied the value for only one map (the one that had the maximal value that was still below 255, or the minimal value above 0) and did the same for the other two maps. This was done only based on the minimal/maximal value of one map so the color combinations won't change in the combined image of all 3. If any of the 3 maps was already at 255 then no change was made. We named the image that now covers all the available color range the 'normal' image.

In order to quantitatively claim that no images were either over or under exposed we later tested the images. Each image was separated to its 3 color channels (Red, Green, Blue). For each channel we computed the color histogram (8 bits of available colors per color channel, see example in Figure 36).



Source image



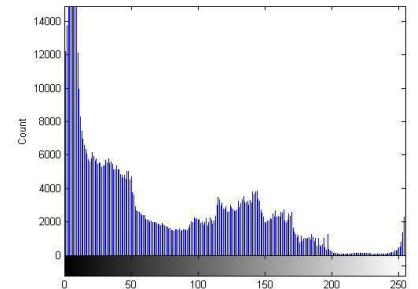
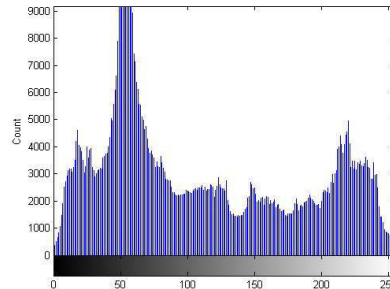
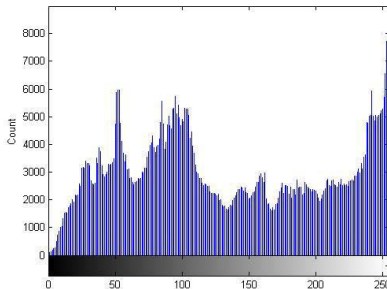
Red channel



Green channel



Blue channel



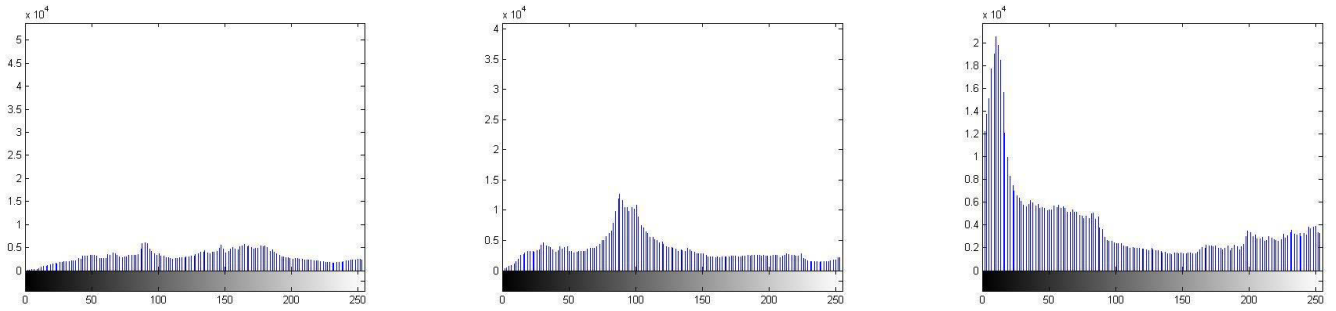


Figure 36 - The exposure normalization process

Source image (**top row**) is broken into its 3 color channels (**second row**): red, green, blue. Histogram of the colors in each channel (**third row**) is generated — ranging from black (0) to white (255). The histogram for each channel separately is then normalized (**bottom row**) such that the color value are distributed normally within the range by linear shifting of the colors. After the normalization is conducted, the 3 channels are re-combined to create the normalized exposure image.

Eye tracking

Eye-position data were acquired at 1000 Hz using an EYELINK-1000 (SR Research, Osgoode, Canada) eye-tracking device. The images were presented on a CRT2 screen (120 Hz), using Matlab's psychophysics and eyelink toolbox extensions (Cornelissen, Peters, & Palmer, 2002). Stimulus luminance was linear in pixel values. The distance between the screen and the subject was 80 cm, giving a total visual angle for each image of $28^\circ \times 21^\circ$. Subjects used a chin-rest to stabilize their head. Data were acquired from the right eye for most subjects. 4 subjects who could not calibrate with the right eye used the left eye. All subjects had normal or corrected-to-normal eyesight.

Modeling using the saliency model

Commonalities between different individuals' fixation patterns allow computational models to predict where people look, and in which order. There are several models for predicting observers' fixations, some of which are inspired by putative neural mechanisms. A frequently referenced model for fixation prediction is the Itti et al. saliency map model (SM) (Itti, Koch, & Niebur, 1998). This "bottom-up"-based model is based on contrasts of intrinsic images features such as color, orientation, intensity, flicker, motion, and so on, without any explicit information about higher-order scene structure, semantics, context, or task-related ("top-down") factors, which may be crucial for attentional allocation. Such a bottom-up saliency model works well when higher-order semantics are reflected in low-level features (as is often the case for isolated objects, and even for reasonably cluttered scenes), but tends to fail if other factors dominate: e.g., in search tasks, strong contextual effects, or in free-viewing of images without clearly isolated objects, such as forest scenes or foliage. For my work I used the C++ version of the saliency model maintained and implemented by the IILab at USC. The parameters used were the standard replicable ones of the model, which means the parameter *-norand* was always present to allow for repeated accurate results. This implies that the model doesn't make full use of its capabilities as it does not incorporate any statistical methods, but returns reliable results. This was shown to not affect the results significantly. The normalization used was always *maxnorm*. Other than that, the parameters were used differently in each study and are reported in each chapter separately. Illustrations of the saliency model and its usage in each dataset are shown in the relevant chapters as well.

Single-neuron recordings

Spike detection

Single unit activity is recorded via bundles of microwires connected to a preamplifier module, providing a gain of 50,000 over a bandpass of 0.3 Hz to 6 KHz.

The currently used algorithm to extract spikes from the raw data can be summarized as follows:

1. Apply a high pass filter (typically, 300–1000 Hz) and obtain H .
2. Compute the median M of $\text{abs}(H)$, and determine a fixed threshold $T = \frac{M}{0.675 \times \text{std}(\text{baseline})}$.
3. Find points that satisfy $H \geq T$.
4. Iterate over the detected points. Take a short time interval (64 samples) and find the maximal amplitude.
5. Align the short interval according to the maximal point. Throw away all points that occur within the refractory period (3 ms).
6. Use clustering method (wavelet based) on the 64 samples to determine neuron clusters.

This method generally finds clusters with different spike shapes. However, it can also have many false alarms for channels without many spikes. One way to discard all these is to look at

the inter-spike-interval (ISI), and at the total number of the spikes. This was done manually post-analysis to the channels that were shown to be selective.

LFP Analysis

60 Hz noise removal

A recurring theme in processing this signal is that a strong 60 Hz noise appears in the raw data. This is due to the AC power supply. The noise can also appear in higher octave bands (typically, at multiples of 60 Hz). The straightforward approach to solve this problem is to apply a notch filter (also known as band stop filter) at 60 Hz. The ideal filter stops only 60 Hz, but typically, people would apply a band stop filter that starts attenuating the signal at 59 Hz, reaching maximal stop at 60 Hz and going back up until reaching 61 Hz. The simplest way to implement this filter is using Matlab's *iinotch* function, which designs a second-order IIR filter. Second order means that this filter has two components (called a, b). The raw signal is first convolved with kernel a. It is then reversed and filtered with kernel b. Finally it is reversed again. IIR means "infinite impulse response".

Low-pass filtering

The LFP signal refers to low frequency bands, typically in the range of 0–130 Hz. Some people use the notion of low/hi-LFP (0–60 Hz, and 60–130 Hz, respectively). To obtain LFP signal, most people use a so called “second-order FIR elliptical filter”. The filter has two components (b,a), similar to the IIR described above. The easiest way to construct this filter is using Matlab's *ellip* function.

After filtering the signal and obtaining the low-pass version, it is possible to re-sample according to the new Nyquist frequency. i.e., if we low pass the signal at 100 Hz, we can sample the signal at 200 Hz. Given the original sampling frequency f_s , this corresponds to sampling the new signal every $\frac{f_s}{200}$ seconds. This significantly reduces the amount of data that needs to be stored for further analysis.

Experiment presentation

As all studies with the epilepsy patients were done at the clinical setup at UCLA is it important to pay attention to the environments and conditions by which they happened to give the reader a view of the situation.

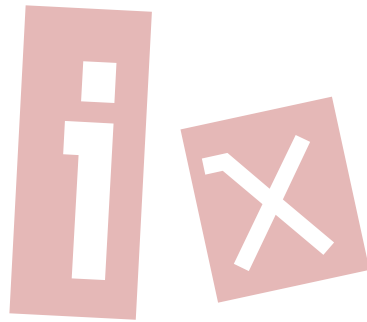
Patients usually lie on a bed in a room by themselves or with their close relatives. During any given experiment the room included myself and at least the lab research assistant who was in charge of interacting with the patient in cases where he might feel unwell, or might want to stop the experiment but might feel uncomfortable saying so to the “eager” scientist. The research assistant was also in charge of setting the electrode connections from the pre-amplifiers on the patients’ brain to the acquisition system. Before each experiment we would tell the assistant which of the electrodes we wanted to include in our 64 channels recorded (out of, sometimes, more than 8 x 8 available brain regions). Gains were manually set on the acquisition system to allow for the recording to yield higher signal-to-noise ratio. The recordings for my studies were done on a digital “*Cheetah*” that would send the raw data both to a hard drive where it will be recorded, and also via TCP/IP to a powerful speedy, dedicated computer connected through the internal network. This computer would run a simple server that would collect the data and detect the spikes. This computer, in turn, will also act as a client for the experimental computer,

which would receive from it the firing rates based on the spikes detected. This process of acquisition, recording, spikes detection, and data transfer takes approximately 17 ms for a single channel, and about 80 ms for 4 channels. The system would re-sync itself every 100 ms and would make sure the 3 computers were transferring data correctly.

The patient himself would work on the final, experimental computer, which typically would be my 15' HP Pavilion dv6000 laptop running Matlab and the psychophysics toolbox for accurate image timing and presentation. Sync pulses between the experiment computer and the acquisition systems would be sent by Matlab using an ActiveWire USB device.

Internal Review Board and HIPAA

All studies at Caltech and UCLA were approved by the Caltech and UCLA Internal Review Boards. Patients signed an informed consent prior to each study. Autism and AgCC subjects also signed an additional consent form before attending a clinical psychological interview with Dr. Lynn Kerlin Paul. I attended 4 of the interviews and conducted one interview with the amygdala-lesion patient SM, and 2 interviews with the prosopagnosia subjects MH and KM. These interviews were fully recorded and archived as MP3 files. All clinical data and records of studies are kept in a log book at the Klab at Caltech. All patient data is kept at the UCLA medical center. No internal patient or subject data is kept on any device but an external hard drive in the possession of the labs at UCLA and Caltech. During my years at Caltech I attended online classes pertaining to handling human patient data twice and received the NIH certificate for each.



Observers Are Consistent When Rating Image Conspicuity

*You are very unique;
Just like everybody else.*

T-shirt logo



⁶ person walks in a museum. He looks at the artwork, the sculptures, the statues, the Mona Lisa, the sunflowers, the scream, some Picasso, Chagall. He likes this painting a lot, doesn't feel strongly about that one. He is inspired by the colors and the emotions expressed in the small one just across from the hallway, spends a few moments next to another one, and really feels he possess a unique taste in arts. Only his set of beliefs, emotions, history, memories, and background — his complex personality — led him to develop such a unique taste. He really loves the sunflower and really doesn't like the Monet. He is special. He has his own style.

Well, it turns out that it is not necessarily so. Sometimes, independent observers — who share little in their emotions, history, background, beliefs, hopes and psychology — will end up exhibiting a much more common flavor upon their stroll through the museum. Much more conventional and similar to others than they would like to think. You are very unique, just like everyone else.

The question of the consistency across observers when witnessing the same piece of art, and questions related to consistency with oneself were the basis of the very first study I conducted at Caltech. A seemingly trivial question that my art teacher in high-school seemed to have had a trivial answer to from an early stage ended up with some surprising and exciting results.

⁶ This chapter is partially based on: Cerf M, Cleary DR, Peters RJ, Einhaeuser W, Koch C, "Observers are consistent when rating image conspicuity", *Vision Research*, Volume 47, Issue 27, Pages 3052-3060, 2007.

Art

Upon my time at Caltech I tried to take as many art classes as I could at the Art Center College of Design, which has a form of collaboration with Caltech. In one of the drawing classes a teacher by the name of David Tal coined the immortal phrase that ended up inspiring this study: “every artist has a limit to what he is capable of doing. He strives to reach that limit, and when he realizes he can’t do better than that – he names it his ‘style’”. And apparently every artist has a unique style. Something that defines him.

The museum stroll

People commonly rank and equate things. You go into a bar and as soon as you set foot in it, you scan the other visitors and put a numerical quantity on each and every one of the girls who are sitting there. The blond one is a 7, the red-head is a 6.3, the tall girl is a 7.1 – a bit cuter than the blond, but not nearly as much as the 8.2 dark-skinned girl at the counter. The bartender is a 5, and so forth. Our brain is an economic entity that rapidly assigns monetary, quantitative, or other numerical values to everything. As said once by Don Corleone in the *Godfather* – everything has a price, and all you need to do is make someone an offer they can’t refuse. When we go into the museum, somewhat transparently to us, we assign a numeric value to each piece of art we observe. We walk in the Louvre and enter the ancient Egypt gallery. We start looking at the artworks and simultaneously start assigning values to them, say the scale goes from 1-9 where 1 is the “ugliest” for lack of an other word, and 9 is the “prettiest”, we constantly give values to every piece of art we witness.

Experiment

In the following experiment we had subjects look at a selected sequence of images from different categories. Each image was presented on the screen for 2 seconds, after which subjects were asked to rate the image on a scale of 1-9. We tried to be as vague as possible with respect to the instructions so we would not bias the subjects to any particular parameter (telling them for instance to rate the images based on how “interesting” they are would bias them in one direction, while telling them to rate the images based on how “pretty” they are would bias them in another). We instructed subjects to tell us how “conspicuous” images are, and were careful to not tell them too much about “conspicuity” – leaving them somewhat baffled by the unclear term so they would choose their own metric as much as possible.

The images were divided into few categories, based on various parameters we felt might be important for their judgment. Mainly, the categories ranged from very abstract images to very semantic. Subjects rated 200 images in a block. The experiment had 5 blocks, in random order, each having a single category. The categories were: 1) fractals (a very abstract set of images, with no clear difference between many of them, as they might only include a tiny variant in color, or orientation of the Lorenz attractor); 2) satellite images (even more so than the fractals, this set of images looks very similar. The satellite images were taken with high resolution but are of a very large ground area, with few or no clear landmarks, making it seem like a repeated Rorschach pattern with only tiny variations); 3) grayscale natural scene images (taken from the Van Hateren database of scenes containing houses, foliage, trees, grass, roads, lakes, plants, etc); 4) color natural scenes (similar to the Van Hateren database, these are images that show nature – views of mountains, canyons, lakes, waterfalls, plants, and static

street and housing scenes); 5) magazine covers (semantic magazine covers of familiar – Time, Sports Illustrated, etc – magazines, combined with unknown – Japanese news, etc – magazines addressing various topics such as fashion, sports, global and local news, health, etc, suggesting more semantic content).

In random block order, subjects took the 200 images sequence in a psychophysics experiment environment and rated each image rapidly after its viewing. In each block, each of the 200 images was repeated 3 times, thus giving a 600 images block. The order was randomized within each block for each subject. Subjects were not told that the images are repeated and were not instructed to do anything in particular given these repetitions.

Scale

The scale chosen for this and all following experiments was 1-9. Such a scale allows tiny variations between the answers (unlike a scale of 1-3 which basically conforms to a stronger claim about each image, a scale of 1-9 allows for more diversity, and thus makes it harder to be accurate in correlating the results later on – more options, higher chance of variability in the results). The scale was also easy to use keyboard-wise as subjects had the entire keypad as their platform, and it's primarily an odd-numbered scale. That is, it allows subjects to choose "5", which is an "uncommitted" answer. A scale of 1-8, for instance would force them to at least slightly commit to one direction (4 – lower than the middle, 5 – higher than the middle) with every choice. We wanted the option to not commit, again, so to allow the subjects an "easier" escape method of no-choice that would "hurt" our statistics if they made use of it commonly, but strengthen the results if they didn't, which indeed was the case.

Outcome

What we saw in the experiment was that people ended up being very consistent in their rating of images. Looking at 200 images in each of 5 blocks, where repetitions occurred, while subjects were not asked to be consistent, most subjects showed a high level of consistency in their rating. More surprising was the fact that their results were highly correlated with those of some other individuals who took the same experiment. Is it possible that our taste in images is much more common than we wish to think? We conducted the same experiment a year later, and saw the same trend. The rating of images are consistent a year apart, even when subjects remember none of the images. To quantify the level of accuracy and to make sure it does not rely on memory alone, we had subjects perform a memory task after the experiment, and we tweaked a bunch of experimental parameters like the order of categories presentation, and the exposure time. The results remained utterly the same. The rating of images by subjects, be they highly semantic and engaging as magazine covers or boring and meaningless as similar patterns of satellite images, show the same general result – when we are asked to give that picture a number, the highest predictor of our rating is not our prior interview, or our psychology as it reflects itself in our preferences, but actually the rating of these very same images by a random observer who we may think we have nothing in common with.

Introduction

Our natural visual environment is dynamic and requires rapid selection of relevant stimuli (Badaruddin et al., 2007; James, 1890). Given the celerity of scene recognition, these processes can therefore be driven to a significant extent by stimulus-dependent factors, rather than by top-down and higher-order cognitive factors (Biederman, 1981; H. Kirchner & S. J. Thorpe, 2006; Li, VanRullen, Koch, & Perona, 2002; Potter, 1976; Potter & Levy, 1969; Potter, Staub, Rado, & O'Connor, 2002; Renninger & Malik, 2004; Rousselet, Fabre-Thorpe, & Thorpe, 2002; S Thorpe, Fize, & Marlot, 1996). Computational studies suggest that object recognition, to some extent, can also be performed in such a sensory-driven (“bottom-up”) manner (Riesenhuber & Poggio, 2000; Sun & Fisher, 2003; S. Thorpe, Delorme, & Van Rullen, 2001). Models of spatial attention often rely on a sensory-driven saliency metric to describe relevant subsets of a stimulus (L. Itti & C. Koch, 2001; C. Koch & Ullman, 1985). Such models predict certain aspects of observers' eye positions, change detection, or pattern of attentional deployment (Deco & Schurmann, 2000; V. Navalpakkam & Itti, 2005; Parkhurst, Law, & Niebur, 2002; Peters, Iyer, Itti, & Koch, 2005; Sun & Fisher, 2003; J. K. Tsotsos et al., 1995; Wright, 2005). However, they typically measure saliency within a static image or video-frame, and do not address the question as to how conspicuous one image is relative to another. A prerequisite for such a measure to exist is that different individuals share common metrics of judging an image's conspicuity. When different observers judge conspicuity, do they form similar metrics and apply them consistently?

Here we ask observers to assign a single measure of conspicuity to images within an image category (e.g., how conspicuous one magazine cover image is relative to other ones; how

conspicuous one outdoor photo is relative to other outdoor photos). In addition, we test recognition memory for these scenes and the effect of presentation duration. Using this experimental setting, we address 3 questions: First, are ratings of image conspicuity consistent across observers? Second, are the conspicuity ratings for the same image consistent across multiple presentations within the same observer? Third, is rating consistency primarily determined by recognition memory, or do low-level stimulus-driven factors play a decisive role? The extent to which these questions have positive answers enables us to construct sensory-driven (bottom-up) models of image conspicuity. Consequently, on a more abstract level our study will provide an upper bound as to how far bottom-up models can capture seemingly subjective stimulus appeal.

Methods

Stimuli

Five sets of images were used in the experiments: (i) colored fractals generated at gnofract4d (<http://gnofract4d.sourceforge.net>) and downloaded from the “Spanky Fractal Database” (<http://spanky.triumf.ca/www/welcome1.html>), (ii) grayscale outdoor photos of trees, shrubs, rivers, and other scenes that did not contain objects like cars, people, or animals (van Hateren & van der Schaaf, 1998), (iii) overhead, grayscale, 10-m resolution satellite images from the NGA database (<http://geoengine.nga.mil/>), (iv) colored natural scenes that included landscapes, flowers, trees, and oceans, and (v) a set of colored contemporary magazine covers (e.g., Time, People Magazine) including text (Figure 37). We chose the images such that – to the best of our possibilities – there are no obvious semantic differences within a category, i.e., images of the same category share about the same gist.

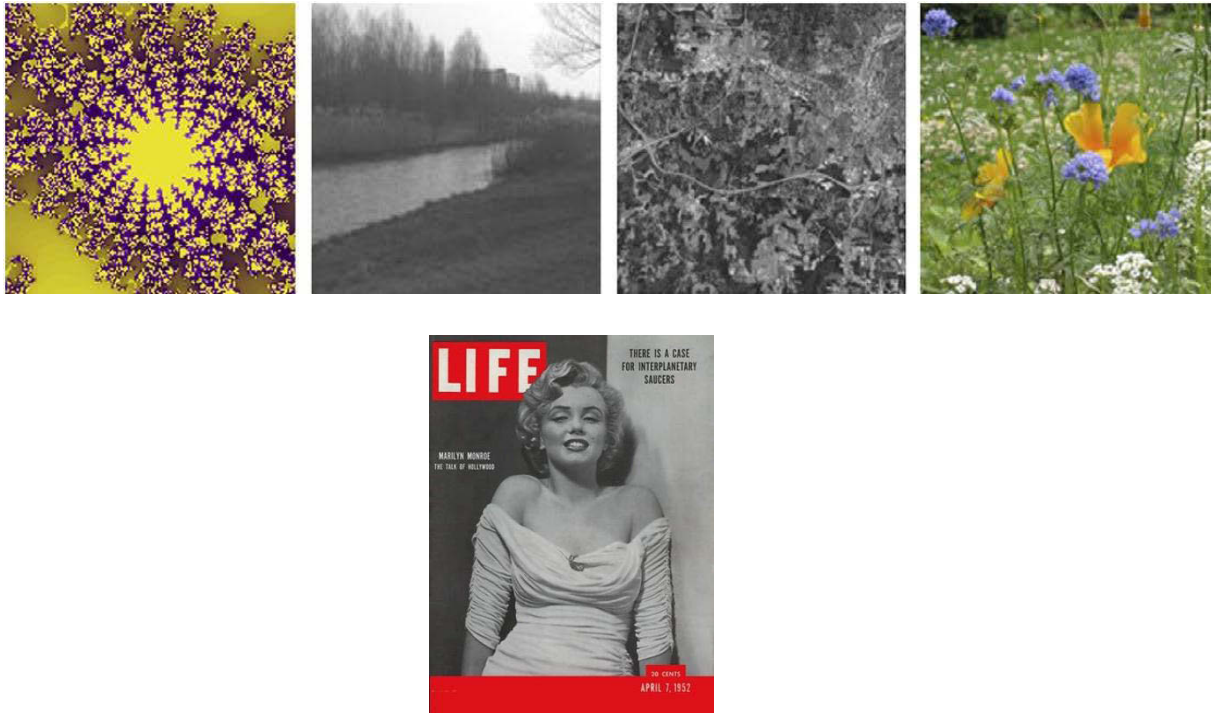


Figure 37 – Image classes used in the study

Observers judged the conspicuity of pictures from 5 different image classes, consisting of colored fractals, grayscale nature scenes, overhead satellite images, colored nature scenes, and contemporary magazine covers.

Participants

Ten volunteers (5 males, 5 females; ages 20 to 41) participated in experiment 1. Four of these 10 participants (2 males, 2 females) also participated in experiment 2. Six additional participants (ages: 21 and 22 males, 20 and 24 females) participated in experiment 3. All observers had uncorrected normal vision and were naïve with respect to the hypotheses tested. All experimental procedures were approved by Caltech’s Institutional Review Board and were performed with the written informed consent of all participants.

Presentation

Participants viewed sets of images on a computer 19" CRT-monitor in a distraction-free, darkened and isolated environment. Participants were instructed to be well-rested for the experiment. To keep a constant distance of 50 cm from the screen, we encouraged participants to use a chin-rest. Using the Groovx framework (<http://ilab.usc.edu/rjpeters/groovx>), we developed an extended Tcl/Tk program that displayed the images at a uniform resolution of 1200 by 900 pixels at 60 Hz.

Conspicuity rating

In the main experiments observers were instructed to rate the conspicuity of images. In the written instructions given at the start of each experimental day, we defined this as “a measure of how “salient”⁷ or noticeable an image is relative to its surroundings” (literal quote from the instructions). We further instructed observers that they would “be asked to determine how salient you find each image relative to the images previously seen in the current set of images.” The instructions furthermore directed observers to only make comparisons within the current image set (and not between sets) and to distribute their responses equally between 1 and 9 (1 being not conspicuous and 9 being very conspicuous). Instructions furthermore discouraged observers from questions on the purpose of the experiment prior to experiment conclusion, but encouraged them to ask any questions needed to clarify experimental procedures.

⁷ Although we used "salient" and "saliency" in the instructions, we refer to it as "conspicuous" and "conspicuity" throughout this chapter to avoid confusion with different notions and definitions of "saliency" in the literature.

Experiment 1

Experiment 1 consisted of 3 separate sessions conducted on different days. In the first two sessions, two blocks of one category each were tested (block design), the third session consisted of a single block. Between the blocks observers took a five-minute break. With the exception of the “magazine cover” category block, which all observers performed in the third session, the order of blocks was randomized across observers (Table 1). To ensure that this relative positioning of the “magazine cover” category had no effect on the results, we tested two additional participants, who had not participated in any of the other experiments, on the “magazine covers” category alone. All data of these two observers were well within the range of the original observers. This result, together with the randomization of the first 4 blocks, renders it unlikely that any of the observed effects is contingent on the order of category presentations.

	Session 1		Session 2		Session 3
	Block 1	Block 2	Block 3	Block 4	Block 5
Subject 1,8	Satellites	Color nature	Fractals	Grayscale nature	Magazines
Subject 2,4	Grayscale nature	Fractals	Color nature	Satellites	Magazines
Subject 3,5,10	Fractals	Grayscale nature	Satellites	Color nature	Magazines
Subject 6	Fractals	Satellites	Grayscale nature	Color nature	Magazines
Subject 7,9	Satellites	Color nature	Grayscale nature	Fractals	Magazines

Table 3 – Order of categories used in experiment 1

Each block consisted of 2 phases: a conspicuity rating phase, and a memory phase. In the first phase observers rated conspicuity following the instructions as described above. At the start of each session, observers saw a screen with a reminder to use the values 1 to 9 for their responses. The experiment began when observers pressed the space bar. For each trial, the image was displayed for 600 ms on the entire screen (42x32 degrees of visual angle), followed by a request for a response and a count of the number of images remaining in the session. This request remained on-screen until the observer responded.

At the start of each block, observers saw 35 training images drawn from the same category so they could develop a consistent internal metric for judging images (Figure 38). The training images did not reappear in the further course of the experiment and were excluded from

analysis. Subsequently, observers viewed and responded to 300 images in three repetitions of 100 unique images from one of the five sets. There was no break between the training images and the entire 300-image sequence. Observers merely saw a sequence of 335 images, of which some repeated thrice (the 100 “test” images). While the same training and test set was used in all observers, the order of images within a set was randomized individually. Observers were not told that images could appear more than once, and were in particular never explicitly instructed or encouraged to respond consistently.

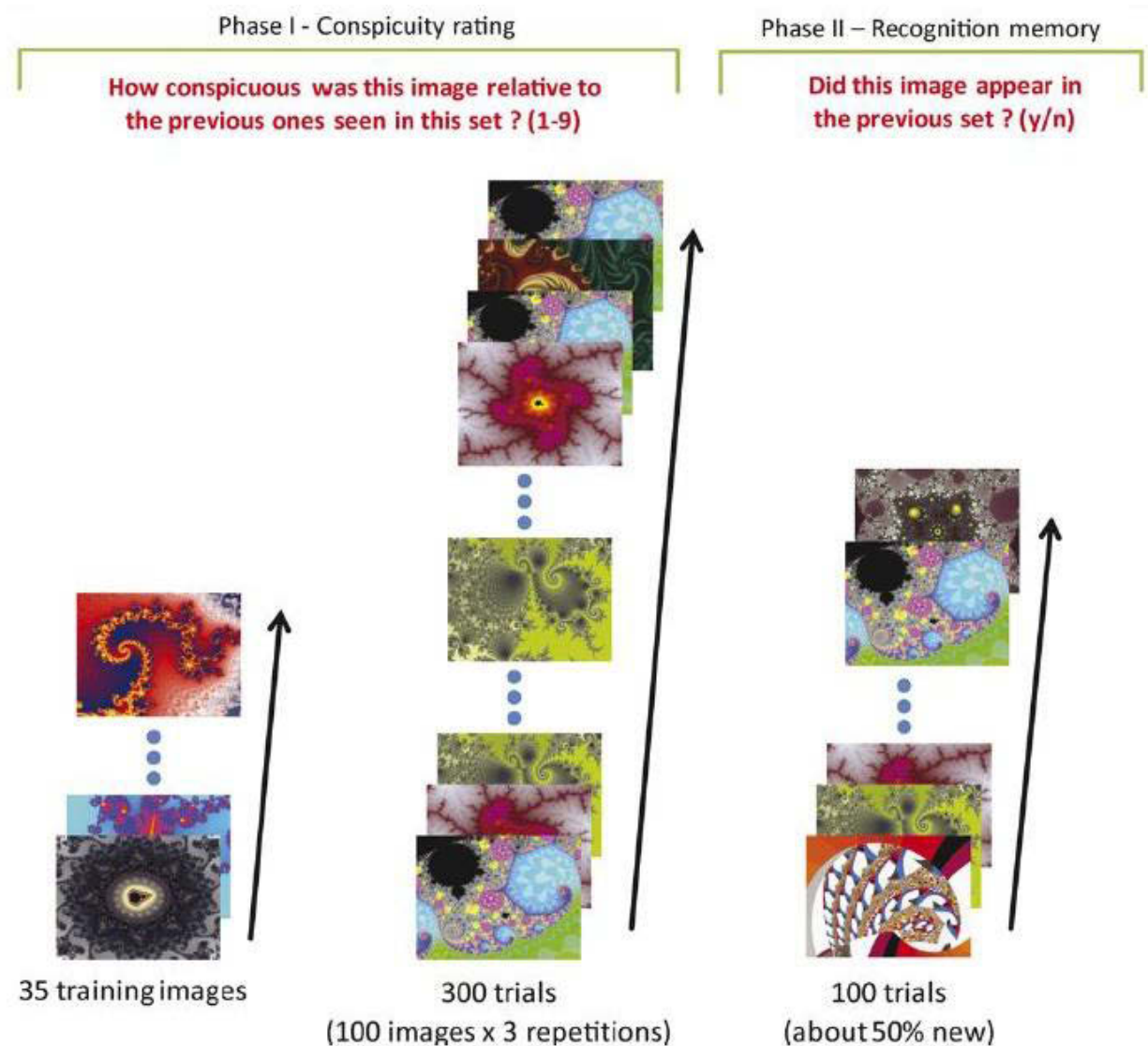


Figure 38 - Conspicuity rating experiment design

A single experimental block as used in experiments 1 and 2. Only one category was tested per block (here: fractals). The first 35 images did not reappear ("training images", left) during the remainder of the block. The next 300 trials (middle) consisted of 100 images that were presented thrice in random order (with the restriction that at least one different image occurred between two presentations of the same image). For all 335 trials, observers rated subjective

conspicuity. Immediately following this “rating” phase, there was a “memory” test phase (right), in which 100 images were presented. About half of the images were repetitions of the rating phase, half were novel images of the same class. Observers responded as to whether or not they had seen the image before. In both phases, images were presented for 600 ms. In experiment 1, each observer performed 5 of these blocks distributed over 3 sessions (Table 3), in experiment 2, four blocks were tested, split across 2 sessions.

In each block, the conspicuity rating phase was followed immediately by a brief memory testing phase. Observers were presented 100 images, about half of which were taken from the previously viewed images, and the remainder novel (the “color natural” and “magazine cover” categories had a 48-52 split between novel and familiar, the remainder was split 50-50). Observers responded if they had previously seen this image or not. Images in the memory task were shuffled randomly, and observers did not know what fraction of the images they had previously seen. The first memory test came as a surprise to the observers, since they had not been specifically instructed to remember the images. This allowed us to test the results of the first block as compared to other blocks in terms of recognition memory and to control for observers’ effort to explicitly remember their responses.

After finishing each session, observers were interviewed about their experience – how interesting they found the different image classes and how they thought they judged conspicuity, along with questions that were aimed at verifying their attentiveness during the experiment and their ability to follow the instructions.

Experiment 2 – Control for the relevance of memory

To test the effect of memory on the conspicuity rating, we repeated experiment 1 with 4 out of the 10 participants about one year after experiment 1. We selected the 4 participants solely on the basis of logistic considerations, i.e., availability on campus, but not based on their results in experiment 1. Participants repeated four of the five image classes using the very same images as in the year before (different order within each class) and otherwise repeated the paradigm of experiment 1. All 4 participants repeated the fractals, grayscale nature, and satellite images categories. Two of the participants also repeated the colored nature images, and the other two repeated the magazine covers. Thus, each participant judged the fractals, the grayscale nature, the satellite images, and one of the two remaining categories (magazine covers or color nature)

Experiment 3 – Effect of presentation duration

In experiment 3, we tested the effect of viewing time, which was kept constant at 600 ms in the previous two experiments, on the conspicuity measure. Six additional observers viewed colored natural images in four blocks of different presentation durations (20 ms, 600 ms, 3 s, and again 20 ms). That is, each of the 100 images was seen 4 times, once in each block. Images were randomly shuffled within each block. Preceding the initial block, participant saw 35 color nature images that were not used in the remaining 100 for 600 ms to form a metric of conspicuity. Otherwise the same settings and instructions were used as in experiment 1.

Results

Experiment 1 – Phase 1 – Conspicuity rating

Intra-observer correlation

Figure 39 depicts the time-course of one observer's (no. 4 – selected at random) judgments for fractal images. Although the responses look random at a first glance, the representation in Figure 39, in which the same data are sorted by image number, shows high consistency across the three presentations of each image. We quantify this consistency by computing the correlation coefficient between the first and second set of responses, the second and third, and the third and first appearance of each individual image. For the example observer of Figure 39 these values were $r = 0.83, 0.90,$ and $0.83,$ respectively.

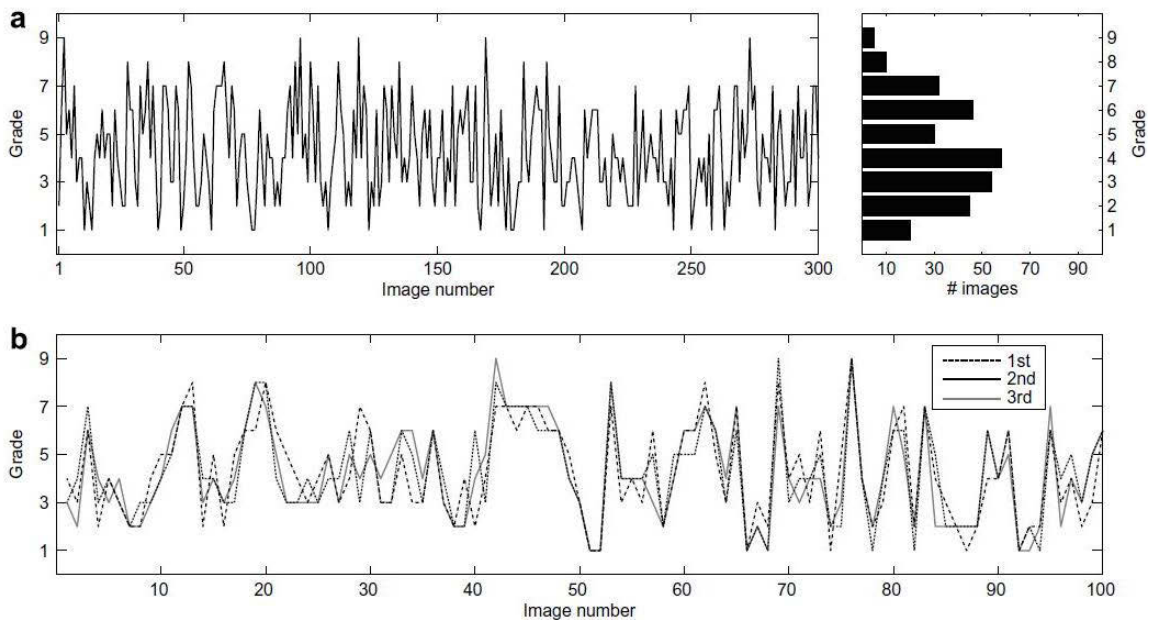


Figure 39 – Observer 4's responses to fractal images

- (a) The conspicuity judgment of observer 4 in the order in which the observer viewed the 300 fractals images. On the right is a histogram of the distribution of the 300 responses.
- (b) When re-arranged by image identity, the consistency of the observer in assigning the same (or similar) conspicuity values to the same image becomes apparent. We found similar trends in 9 out of 10 observers and across all image classes.

To investigate whether the same pattern holds for all individuals, we computed pair-wise correlations for the five image classes for all observers after transforming the conspicuity grading to z-scores (subtracting the mean of the entire set of answers for that observer and that set of responses, and dividing by the standard deviation) to make the responses of all observers comparable. For all but one observer (no. 10), all correlation coefficients exceed 0.4 (Figure 40). These correlations are significantly different from 0 ($p < 10^{-4}$ for any pair-wise correlation). Note that significance prevails at a level of 0.005 even after a conservative Bonferroni correction for the 50 comparisons (as $10^{-4} = 0.005/50$). Observer 10 has notably less consistency than the other ones. His behavior during the experiment was idiosyncratic. It is likely that he was not really following the instructions; thus, his data were excluded from further conspicuity analysis, unless stated otherwise.

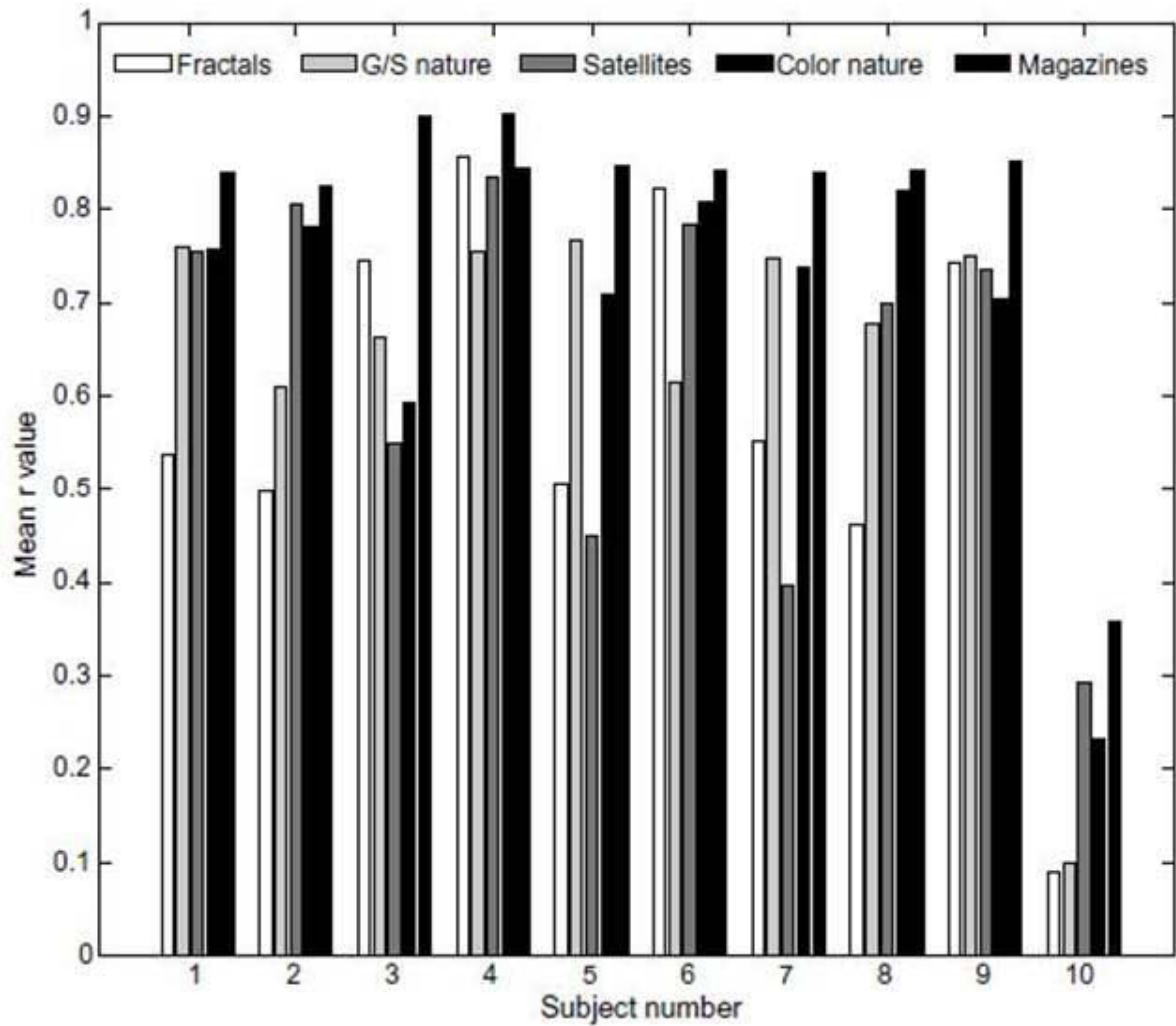


Figure 40 - Intra-observer correlation

Mean pair-wise correlation between first and second answers, second and third, and first and third answers. Observers, with one exception, are consistent when judging the conspicuity of the same image three times. The r correlation values are given as a function of individual and image class.

Within-observer correlation depended significantly on image class (ANOVA, $p = 0.002$, $F[44] = 5.02$), with the mean (over 9 observers) equal to 0.64 ± 0.05 (mean \pm standard error of the mean) for the fractals, 0.70 ± 0.02 for the grayscale natural scenes, 0.67 ± 0.05 for the overhead satellite imagery, 0.76 ± 0.03 for the colored natural scenes, and 0.85 ± 0.01 for the magazine covers. Thus, observers were quite consistent in their conspicuity judgments for the same image, with their consistency highest for the magazine covers and the lowest for the fractals. While it seems surprising that magazine covers, which are of high semantic load, are most consistent, we want to stress that the present experiment was not designed for inter-category comparisons. The fact that we find an effect for any of the categories, irrespective of semantic load, is nonetheless striking and is to be tested in further research.

All of the 9 observers have a higher correlation between their second and third responses as compared to the correlation between their first and second responses (Figure 41), with all but 6 points (out of 45) falling above the diagonal. A sign-test⁸ reveals that this bias is significant ($p = 5 \times 10^5$).

⁵ The sign-test tests against the null hypothesis that both correlation values are drawn from the same, arbitrary (but continuous) distribution and just compares whether r_{12} is larger or smaller than r_{13} irrespective of their absolute values. This is the most conservative estimate; additional assumptions on the distribution would yield lower p-values.

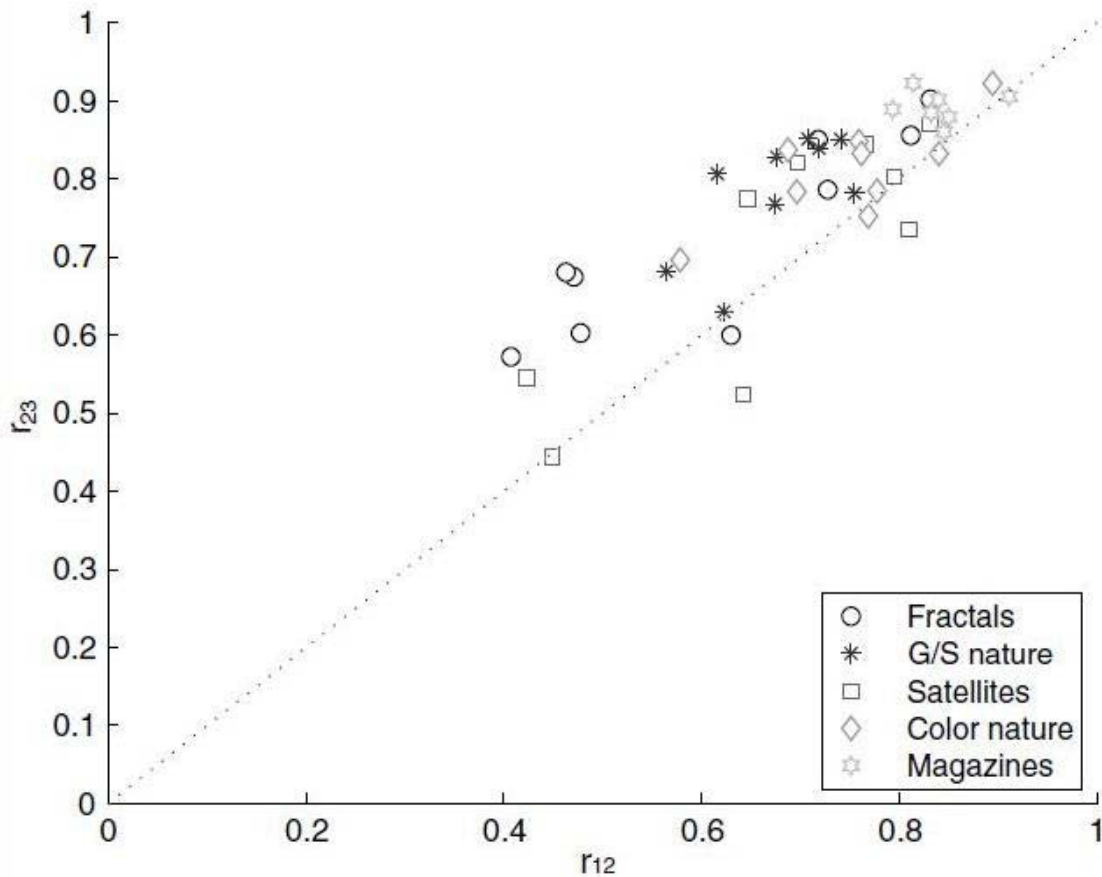


Figure 41 - Observers become more consistent after repeated exposure

Scatter diagram of the r correlation values in the conspicuity judgments between the first and second showing of an image (x-axis) and its second and the third repetition (y axis) for each image class for the 9 observers with high consistency.

As an additional measure beyond linear correlation, we counted the number of images to which a given observer responded identically thrice (e.g., rating image no. 29 a 7-7-7). The example observer (no. 4) responded to 22 fractals with the same rating on all three trials. The average number across the 9 analyzed observers are 12.78 ± 7.17 (mean \pm s.t.d.) for fractals, 15.65 ± 8.30 for grayscale nature, 18.44 ± 6.54 for satellite imagery, 22.33 ± 7.50 for color

nature, and 32.11 ± 11.99 for magazine covers. This is far above the chance level of random occurrence of $100/92 = 1.23$ for each image class, and further supports the high intra-observer consistency reflected in the correlation analysis.

Inter-observer correlation

Analyzing the responses for the 100 images across 9 observers revealed significant correlations of the conspicuity judgments across all observers and categories, though the correlations were lower than the intra-observer correlations (Figure 42). Correlations range from $r = 0.09$ for fractals to $r = 0.51$ for magazine covers. These findings demonstrate that observers are consistent across categories. While we do not deny a category-dependency of this effect (ANOVA, $p < 10^{-9}$, $F[179] = 23.55$), it is intriguing that the correlation is significant for a wide variety of stimuli, ranging from fractals, which have no obvious semantic content, to magazine covers, which seem intuitively largely dominated by content.

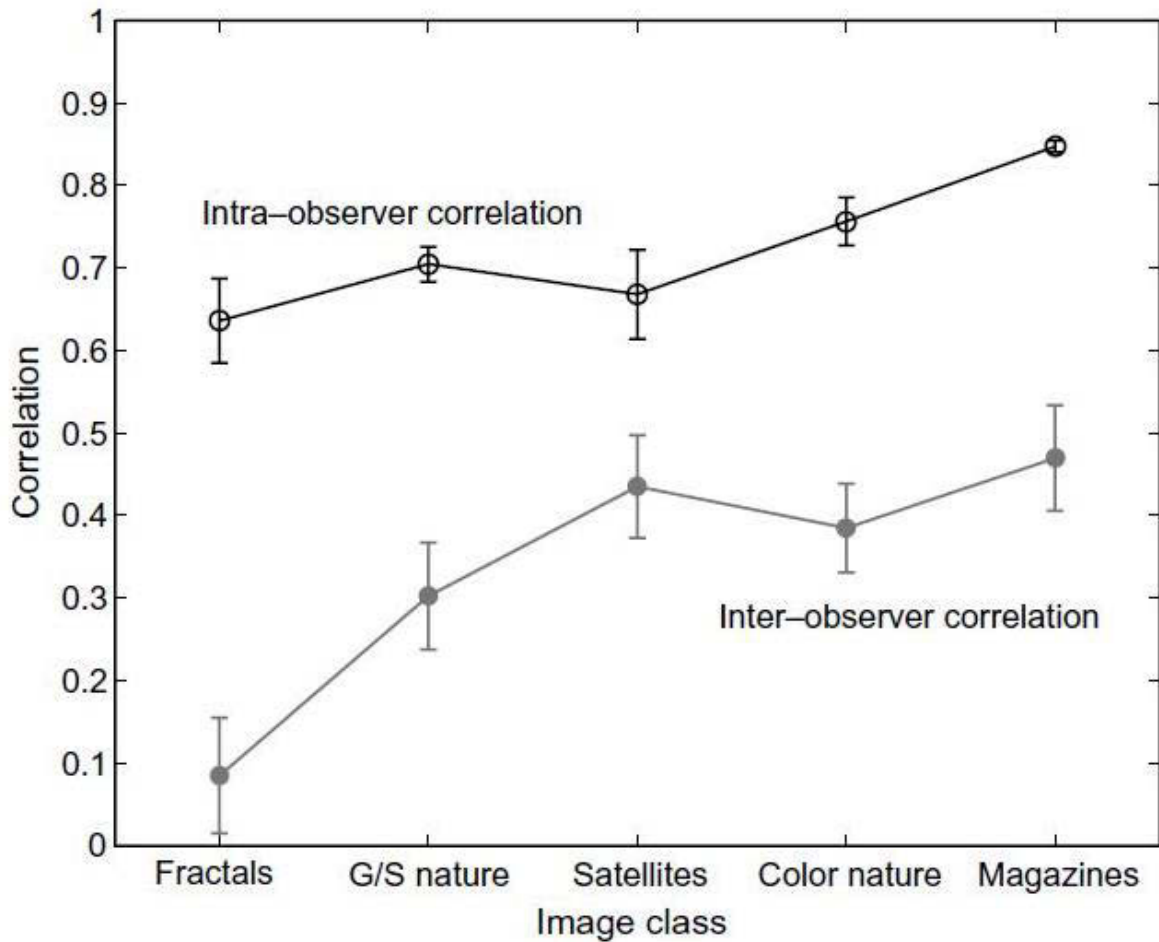


Figure 42 - Inter-observer correlation

Mean pair-wise r-values and standard error of the mean among the 9 consistent observers when judging the conspicuity of identical images as a function of image class. Each r-value is significantly different than 0 at $p < 0.05$. The upper line corresponds to the mean and standard error of the within-observer correlation (Figure 40). Across-observer correlation was calculated for each participant and then averaged across participants.

Experiment 1 – Phase 2 – Memory

How much of an observer's internal consistency can be explained by a sensory-driven conspicuity module that responds in a stereotypical manner to repetitions of the same image and how much has to be attributed to memory for individual images? After each block of 300 images, we asked observers to view an additional set of 100 images, about half of which had been presented during the preceding conspicuity grading phase. All observers performed the memory test above chance (50%) for fractals ($69.9\% \pm 11.5\%$; mean \pm s.t.d. over observers; Figure 43), grayscale images ($90.2\% \pm 6.7\%$), color images ($85.3\% \pm 7.5\%$), and magazine covers ($96.3\% \pm 5.4\%$). Across observers these numbers are significantly different from chance ($p = 3.9 \times 10^{-4}$, $p = 1.5 \times 10^{-8}$, $p = 1.2 \times 10^{-7}$, $p = 5.7 \times 10^{-10}$, t-test, respectively). In contrast, performance was close to – but still significantly above – chance for satellite images ($55.0\% \pm 6.5\%$, $p = 0.04$). This indicates that observers have good recognition memory for briefly presented images of a variety of classes. If memory were the primary factor for consistency we would expect a comparably low intra-observer consistency for the image class that is remembered worst. However, the consistency for satellite images is well within the range of the other classes, which are remembered better. This argues against the notion that rating consistency is primarily memory-driven.

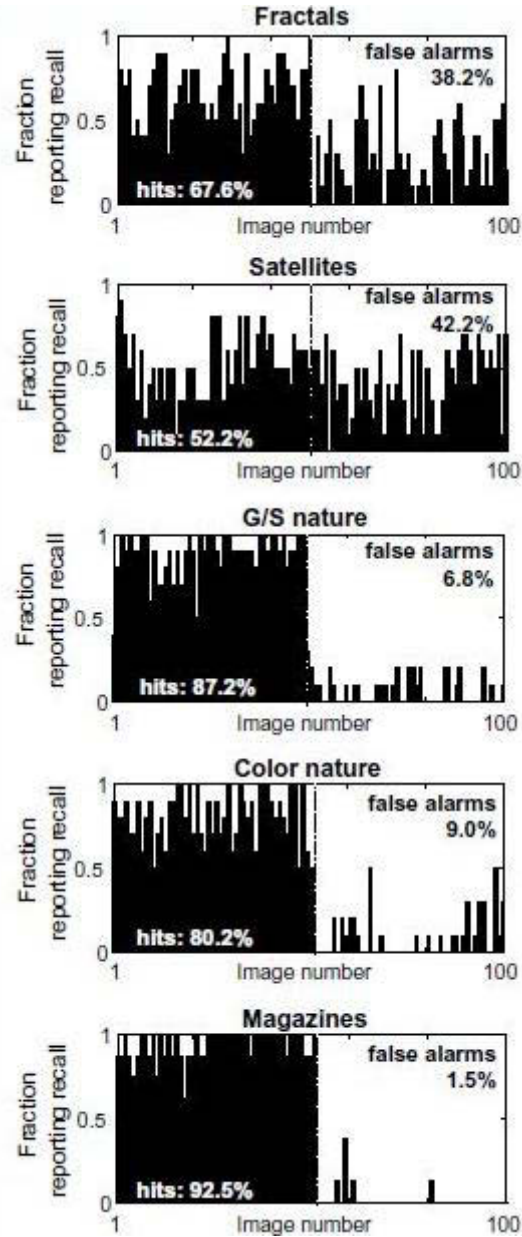


Figure 43 – Recognition memory

Normalized performance on the recognition test for all 10 participants for the 50 previously seen images versus the 50 (52 and 48, respectively, for the color and magazine covers) novel images drawn from the same class. The data are sorted so that recognition performance with perfect memory would be 1 for images 1-50 and 0 for images 51-100 (1-52, and 53-100,

respectively, for the color nature and magazine covers). Each figure includes the hits and false alarms percentages. Observers have almost perfect memory for magazine covers but perform much worse (yet still above chance) for fractal. The satellite images performance is at chance.

Does the rating an observer gives for an image's conspicuity relate to recognition memory? In particular, are highly conspicuous images remembered better? For each observer we selected those images that they rated consistently with the highest score (e.g., responses of "8-8-8", if 8 were the highest value the observer gave for that category). Across the 9 observers, for whom we had obtained consistent conspicuity, we obtain thus 58 "high conspicuity" images (out of $4500 = 9 \text{ observers} \times 100 \text{ images} \times 5 \text{ categories}$), of which 26 were probed in the memory tests. Of these 26, all but one (96.2%) were remembered correctly. Similarly we selected the images with the consistent lowest scores (e.g., 1-1-1). This yielded 163 "low conspicuity" images of which 70 were probed in the memory test. Of those 70, 68 (97.1%) were correctly recalled. In comparison, of all 2286 images ($9 \times [3 \times 50 + 2 \times 52]$) that were targets in the memory test (i.e., were rated in the preceding phase), only 1790 (78.3%) were correctly identified as repetition. A χ^2 -test rejects the null-hypothesis that the factors "conspicuity" (low, medium, high) and "memory" (hit/miss) are independent ($p = 1.3 \times 10^{-4}$; $\chi^2 = 17.95$; 2 degrees of freedom). This suggests that repetitions of images with consistently extreme (high or low) conspicuity are more likely to be reported than repetitions of images with intermediate ratings. We have no means of assessing the conspicuity of images that have not been presented before. We therefore cannot know whether images of extreme conspicuity that were previously not shown would have high false alarm rates for recall. Hence, it remains open whether increased recall for extreme conspicuity is due to better memory or lower response threshold during retrieval. In

either case, images of extreme conspicuity are qualitatively different with respect to memory, be it in memorization or retrieval.

Experiment 2 – Long-term longitudinal consistency

A year after experiment 1, four of the ten original observers (nos. 2, 4, 7 and 9) took the experiment for the second time (experiment 2). All four observers reported in debriefing after experiment 2 that they had not recalled any of their conspicuity ratings from the previous year. The mean change for the intra-observer correlation over the image categories was -0.033 for the 4 observers. For logistic reasons, only 3 categories (fractals, grayscale nature, satellites) were taken by all observers, whereas a remaining category was taken by only two observers each (Table 2). Out of 16 (4 observers x 4 categories) correlation coefficients, 14 were larger for the correlation between second and third presentation than between the second and first presentation. This trend is significant ($p = 0.004$, sign-test) and consistent with the data of experiment 1. Comparing the mean r values from experiment 1 for each image class with those of experiment 2 for the 3 categories taken by all observers yields only small changes (change of mean r : fractals +0.05, grayscale nature -0.07, satellite images -0.04). Recognition memory also remains virtually unchanged: the fractals miss percentage in year two is 32% (comparing to 32.40% in year one for 10 observers), grayscale nature is 9% (12.8% in year one), and overhead satellite imagery have 29% misses (47.8% in year one). To further quantify any behavioral change, we compared the correlation (z-score) between the third viewing of the images in year one with the first viewing in year two. The mean intra-observer correlation between the years was 0.51 (with $p < 0.01$ for all individual correlations). Finally, the mean inter-observer correlations for the three images classes taken by all four observers remain unchanged. While

the mean pair-wise inter-observer correlations for those 4 observers alone in the previous year were 0.15 ± 0.13 (fractals, mean \pm standard error), 0.39 ± 0.06 (grayscale nature), and 0.50 ± 0.09 (satellite images), the changes in the following years were: fractals $+0.065$, grayscale nature -0.030 , and satellite images $+0.107$. Interestingly, despite the lack of explicit memory for satellite images, inter-observer consistency increases, pointing to factors common to all observers, i.e., image-immanent, sensory-driven variables, which determine conspicuity ratings. Although observers do not explicitly recall their previous conspicuity ratings, their overall judgments remain highly consistent after a period of one year which hints for an internal metric that observers keep when making judgments, with the same caveats (changes with time between first to second/third observations), and the same variance with respect to other people.

	Session 1		Session 2	
	Block 1	Block 2	Block 3	Block 4
Subject 2,7	Fractals	Grayscale nature	Satellites	Magazines
Subject 4,9	Fractals	Grayscale nature	Satellites	Color nature

Table 4 - Order of categories used in experiment 2

Experiment 3 – Effect of presentation duration

In experiments 1 and 2, each image was presented for 600 ms. Does this choice of presentation duration affect inter- and/or intra-observer consistency? We presented a set of 100 colored natural scenes to 6 additional observers, none of whom had participated in experiment 1 or 2. Images were shown at 3 different presentation durations in 4 blocks, for 20 ms, 600 ms, 3 s and, again 20 ms. Each block consisted of the same 100 images in random order, (i.e., there were 4 blocks with 4 presentation durations in total). All pair-wise correlation coefficients were positive for all presentation durations. The mean correlation coefficient across all 15 pairs of observers was significantly different from 0 for all presentation durations ($p < 10^{-4}$, t-test for any duration). It is lowest for 20 ms presentations (first block: 0.17 ± 0.10 - mean \pm s.t.d., fourth block: 0.23 ± 0.16), intermediate for 600 ms presentation (0.26 ± 0.14) and highest for 3000 ms (0.33 ± 0.15). An ANOVA analysis for factor presentation time (for the first three blocks) shows a significant effect at $p = 0.0068$. That is, within-observer consistency increases with prolonged viewing time.

Discussion

The present study demonstrates that judgments of an image's overall conspicuity are consistent across observers and across multiple presentations within the same observer. Despite some dependence on image class, a significant correlation between conspicuity ratings is observed for a wide variety of classes, ranging in semantic content from fractals to magazine covers. Consistency within an image class does not depend on memory for images in this class.

In principle, the observed consistency can arise from a variety of sources: one possibility that explains both within- and between-observer consistency is that conspicuity judgments are primarily sensory-driven, image-immanent. This view is supported by the relatively high consistency that already exists for very short presentation durations of 20 ms. An alternative explanation for within-observer consistency would be that observers rely on memory rather than on a stereotyped bottom-up evaluation (Standing, 1973). While observers have good recognition memory for most of our images, such an account would not explain between-observer consistency. Furthermore, in the long-term test after a year, we found that observers did not explicitly remember the images and required time to once again set their metric (revealed by increase of intra-observer correlation over each session). Nevertheless, their ratings were consistent with the ratings made in the previous year. Finally, images with recognizable content and objects are more easily remembered than images with less readily identifiable objects (Underwood, Foulsham, van Loon, & Underwood, 2005). In our case, recognition memory for overhead satellite images was worst of all categories, while memory for fractals was better. Nevertheless, satellite images were rated more consistently than fractals, further arguing against a purely memory-based account.

Although a large body of psychophysical, electrophysiological, and computational studies focus on the rapid recognition of the main content, or “gist” of a scene (Biederman, 1981; Evans & Treisman, 2005; Li et al., 2002; Aude Oliva & Torralba, 2006; Potter & Levy, 1969; Renninger & Malik, 2004; Rousselet et al., 2002; Schyns & Oliva, 1997; S Thorpe et al., 1996), the rapid evaluation of the conspicuity of scenes has received little investigation. We instructed observers to rate the “saliency” of a scene. The rationale behind the choice of such a vague term to characterize conspicuity was to evoke very different associations between individuals as compared to well-defined terminologies. In light of this vagueness, the high consistencies we found are more convincing than if we had used a much more detailed and explicit instruction or pre-labeled example images. In the modeling literature, “saliency” is typically used in the context of objects within an image rather than across images, i.e., as a term of spatial attention rather than global preference (L. Itti & C. Koch, 2001; C. Koch & Ullman, 1985). Wright (Wright, 2005) demonstrated that such within-image saliency is closely linked to the probability of change detection and a subjective saliency rating. Since there is evidence that visual processing of individual objects recruits different early mechanisms from gist recognition in entire scenes (Aude Oliva & Torralba, 2006), our findings on image-wide consistency are consistent, but distinct from Wright’s results. Nevertheless, inter-observer consistency is a prerequisite to construct sensory-driven (bottom-up) models that predict non-idiosyncratic aspects of human behavior. As neurally inspired bottom-up models have been successfully applied to spatial attention (L. Itti & C. Koch, 2001; Peters et al., 2005; J. K. Tsotsos et al., 1995), object recognition (Riesenhuber & Poggio, 2000) and gist recognition (Aude Oliva & Torralba, 2006; Renninger & Malik, 2004), our results spur the possibility that such sensory-driven models could also partially predict human preference for certain scenes. Such a putative

model would have a variety of applications, ranging from art to advertisements to human-factors usability design.

Shirt

Subjects who performed well on the task on a scale that is known only to the author of this work, and is based on friendliness, pure bribery, or looks, received a gift on top of their payment for participation, which was a t-shirt made specifically for this task in one of my early Art Center classes. Figure 44, Figure 45 show some of the subjects and two of the collaborating authors with the “saliency experiment t-shirt”.

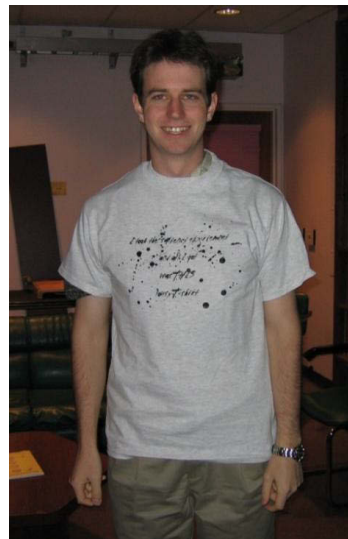


Figure 44 - The saliency experiment t-shirt worn by our subjects



Figure 4.5 - The saliency experiment t-shirt worn by our experimental team

The shirt reads: "I took the saliency experiment, and all I got was this lousy t-shirt".



Attention and High-Level Cues.

Part I – Faces

*“Are we to judge what's on the face,
what's inside the face, or what's
behind it?”*

Pablo Picasso



⁹ Although understanding attention is interesting purely from a scientific perspective, there are numerous applications in engineering, marketing and even art that can benefit from the understanding of both attention *per se*, and the allocation of resources for attention and eye movements. One accessible correlate of human attention is the fixation pattern in scanpaths (Rizzolatti, Riggio, Dascola, & Umilta, 1987), which has long been of interest to the vision community (Buswell, 1935). Task-related (“top-down”) factors, may be crucial for attention allocation (Yarbus, 1967). Here, we test how images containing faces – ecologically highly relevant objects – influence variability of scanpaths across subjects. Although there is an ongoing debate regarding the exact mechanisms which underlie face detection, there is no argument that a normal subject (in contrast to autistic patients) will not interpret a face purely as a reddish blob with four lines, but as a much more significant entity (Hershler & Hochstein, 2005; VanRullen, 2006). In fact, there is mounting evidence of infants' preference for face-like patterns before they can even consciously perceive the category of faces (Simion & Shimojo, 2006), which is crucial for emotion and social processing (R Adolphs, 2002; Barton, 2003).

The contributions of this study is in showing experimental evidence suggesting that subjects exhibit significantly less variable scanpaths when viewing natural images containing faces, marked by a strong tendency to fixate on faces early.

⁹ This chapter is partially based on: Cerf M, Harel J, Einhaeuser W, Koch C, “Predicting human gaze using low-level saliency combined with face detection”, *Advances in Neural Information Processing Systems (NIPS)*, Vol. 20, Pages 241-248, 2007.

Methods

Experimental procedures

Seven subjects viewed a set of 250 images (1024 x 768 pixels) in a three-phase experiment. 200 of the images included frontal faces of various people; 50 images contained no faces but were otherwise identical, allowing a comparison of viewing a particular scene with and without a face. In the first (“free-viewing”) phase of the experiment, 200 of these images (the same subset for each subject) were presented to subjects for 2 seconds, after which they were instructed to answer “How interesting was the image?” using a scale of 1-9 (9 being the most interesting). Subjects were not instructed to look at anything in particular; their only task was to rate the entire image.

In the second (“search”) phase, subjects viewed another 200 image subset in the same setup, only this time they were initially presented with a probe image (either a face, or an object in the scene: banana, cell phone, toy car, etc.) for 600 ms, after which one of the 200 images appeared for 2 s. They were then asked to indicate whether that image contained the probe. Half of the trials had the target probe present. In half of those the probe was a face. Early studies suggest that there should be a difference between free-viewing of a scene and task-dependent viewing of it (Dickinson, Christensen, Tsotsos, & Olofsson, 1994; Henderson et al., 2007; V Navalpakkam & Itti, 2007). We used the second task to test if there were any differences in the fixation orders and viewing patterns between free-viewing and task-dependent viewing of images with faces. In the third phase, subjects performed a 100-image

recognition memory task where they had to answer with y/n whether they had seen the image before. 50 of the images were taken from the experimental set and 50 were new.

Subjects' mean performance was 97.5% correct, verifying that they were indeed alert during the experiment.

The images were introduced as “regular images that one can expect to find in an everyday personal photo album”. Scenes were indoors and outdoors still images (see examples in Figure 46). Images included faces in various skin colors, age groups, and positions (no image had the face at the center as this was the starting fixation location in all trials). A few images had face-like objects (see balloon in Figure 46, panel 3), animal faces, and objects that had irregular faces in them (masks, the Egyptian sphinx face, etc.). Faces also varied in size (percentage of the entire image). The average face was $5\% \pm 1\%$ (mean \pm s.t.d.) of the entire image – between 1° to 5° of the visual field; we also varied the number of faces in the image between 1-6, with a mean of 1.1 ± 0.48 . Image order was randomized throughout, and subjects were naïve to the purpose of the experiment. Subjects fixated on a cross in the center before each image onset. Eye-position data was acquired at 1000 Hz using an Eyelink-1000 (SR Research, Osgood, Canada) eye-tracking device. The images were presented on a CRT screen (120 Hz), using Matlab's Psychophysics and eyelink toolbox extensions (Brainard, 1996; Cornelissen et al., 2002). Stimulus luminance was linear in pixel values. The distance between the screen and the subject was 80 cm, giving a total visual angle for each image of $28^\circ \times 21^\circ$. Subjects used a chin-rest to stabilize their head. Data were acquired from the right eye alone. All subjects had uncorrected normal eyesight.



Figure 46 - Examples of stimuli during the “free-viewing” phase

Notice that faces have neutral expressions. Upper 3 panels include scanpaths of one individual. The red triangle marks the first and the red square the last fixation, the yellow line the scanpath, and the red circles the subsequent fixations. Lower panels show scanpaths of all 7 subjects. The trend of visiting the faces first – typically within the first or second fixation – is evident. (All images are available at <http://www.klab.caltech.edu/~moran/db/faces/>)

Results

Psychophysical results

To evaluate the results of the 7 subjects' viewing of the images, we manually defined minimally sized rectangular regions-of-interest (ROIs) around each face in the entire image collection. We first assessed, in the “free-viewing” phase, how many of the first fixations went to a face, how many of the second, third fixations and so forth. In 972 out of the 1050 (7 subjects x 150 images with faces) trials (92.6%), the subject fixated on a face at least once. In 645/1050 (61.4%) trials, a face was fixated on within the first fixation, and of the remaining 405 trials, a face was fixated on in the second fixation in 71.1% (288/405), i.e., after two fixations a face was fixated on in 88.9% (933/1050) of trials (Figure 47). Given that the face ROIs were chosen very conservatively, i.e., fixations just next to a face do not count as fixations on the face), this shows that faces, if present, are typically fixated on within the first two fixations ($327 \text{ ms} \pm 95 \text{ ms}$ on average). Furthermore, in addition to finding early fixations on faces, we found that inter-subject scanpath consistency on images with faces was higher. For the free-viewing task, the mean minimum distance to another's subject's fixation (averaged over fixations and subjects) was 29.47 pixels on images with faces, and a greater 34.24 pixels on images without faces (different with $p < 10^{-6}$). We found similar results using a variety of different metrics (ROC, Earth Mover's Distance, Normalized Scanpath Saliency, etc.). To verify that the double spatial bias of photographer and observer (Tatler, Baddeley, & Gilchrist, 2005) did not artificially result in high fractions of early fixations on faces, we compared our results to an unbiased baseline: for each subject, the fraction of fixations from all images which

fell in the ROIs of one particular image. The null hypothesis that we would see the same fraction of first fixations on a face at random is rejected at $p < 10^{-20}$ (t-test).

To test for the hypothesis that face saliency is not due to top-down preference for faces in the absence of other interesting things, we examined the results of the “search” task, in which subjects were presented with a nonface target probe in 50% of the trials. Provided the short amount of time for the search (2 s), subjects should have attempted to tune their internal saliency weights to adjust color, intensity, and orientation optimally for the searched target (V Navalpakkam & Itti, 2007). Nevertheless, subjects still tended to fixate on the faces early. A face was fixated on within the first fixation in 24% of trials, within the first two fixations in 52% of trials, and within the three fixations in 77% of the trials. While this is weaker than in free-viewing, where 88.9% was achieved after just two fixations, the difference from what would be expected for random fixation selection (unbiased baseline as above) is still highly significant ($p < 10^{-8}$).

Overall, we found that in both experimental conditions (“free-viewing” and “search”), faces were powerful attractors of attention, accounting for a strong majority of early fixations when present. This trend allowed us to easily improve standard saliency models, as discussed below.

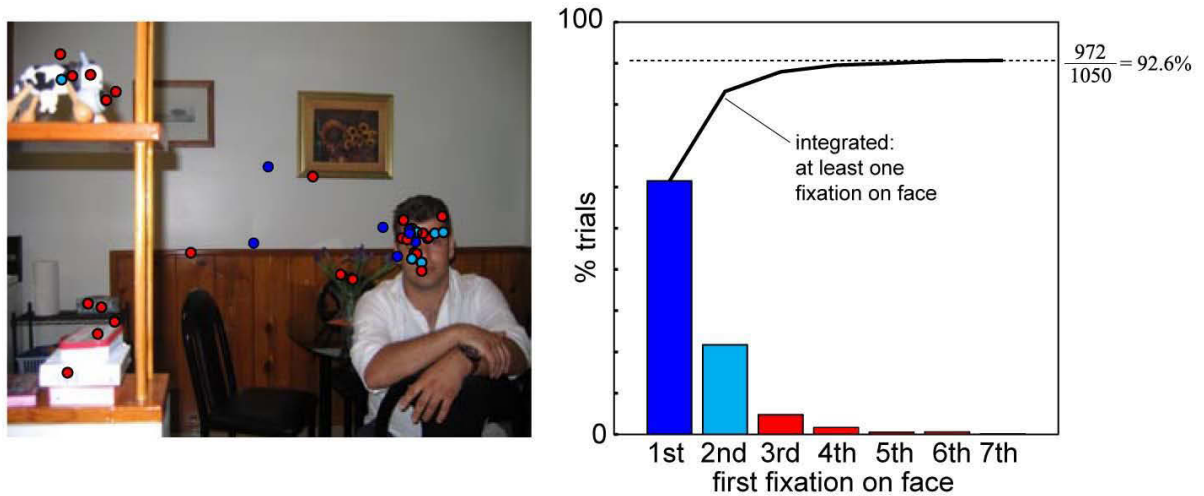


Figure 47 - Extent of fixation on face regions-of-interest (ROIs) during "free-viewing"

Left: image with all fixations (7 subjects) superimposed. First fixation marked in blue, second in cyan, remaining fixations in red. **Right.** Bars depict percentage of trials, which reach a face the first time in the first, second, third, etc., fixation. The solid curve depicts the integral, i.e., the fraction of trials in which faces were fixated on at least once up to and including the nth fixation.

In order to test the importance of faces in the scene we tested the number of faces visited within each image. While most images had only a single face in it, some images had more than a single face. The images that had multiple people in them usually had a variable distance between the faces, including some images where the faces were on opposite sides of the image. Since image presentation is very short (2 s) it is not easy for a subject to visit all the faces in the scene. While the first fixation might go to a face purely because of its attention attractiveness, visiting all the faces is an even stronger effect of attention to faces. The mean number of faces visited pooling all subjects and all images is 94.87% (Figure 48), showing that even when images

have multiple faces and a short presentation time, subjects tend to allocate their resources to visiting all the faces in the image.

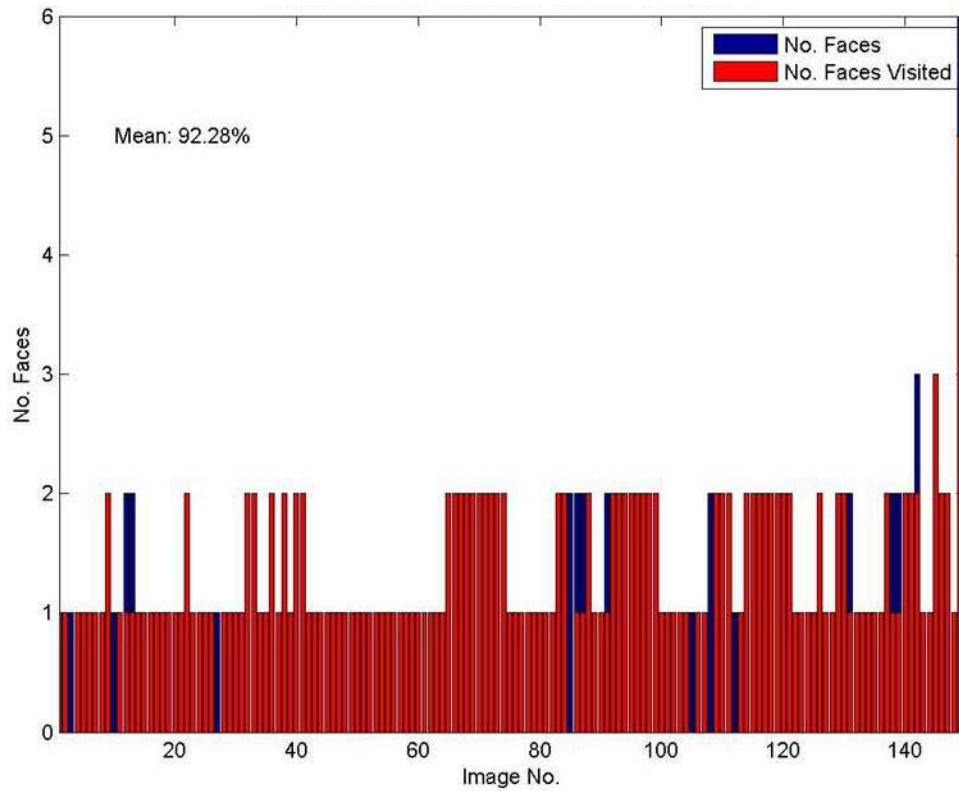


Figure 48 – Number of faces visited for subject 2, experiment 1

Blue bars represent the number of faces in each of 150 images containing faces during the task. Red bars represent number of faces visited by subject 2 in this image. A bar that's only red means that all the faces were visited as the red covers the blue, while a bar that shows blue traces means a face was there but was not visited. Notice image 150 which had 6 faces in it, out of which 5 were visited in the very short presentation time. Subject visited 92.8% of the faces possible in the entire experiment. See results for all subjects in Figure 49.

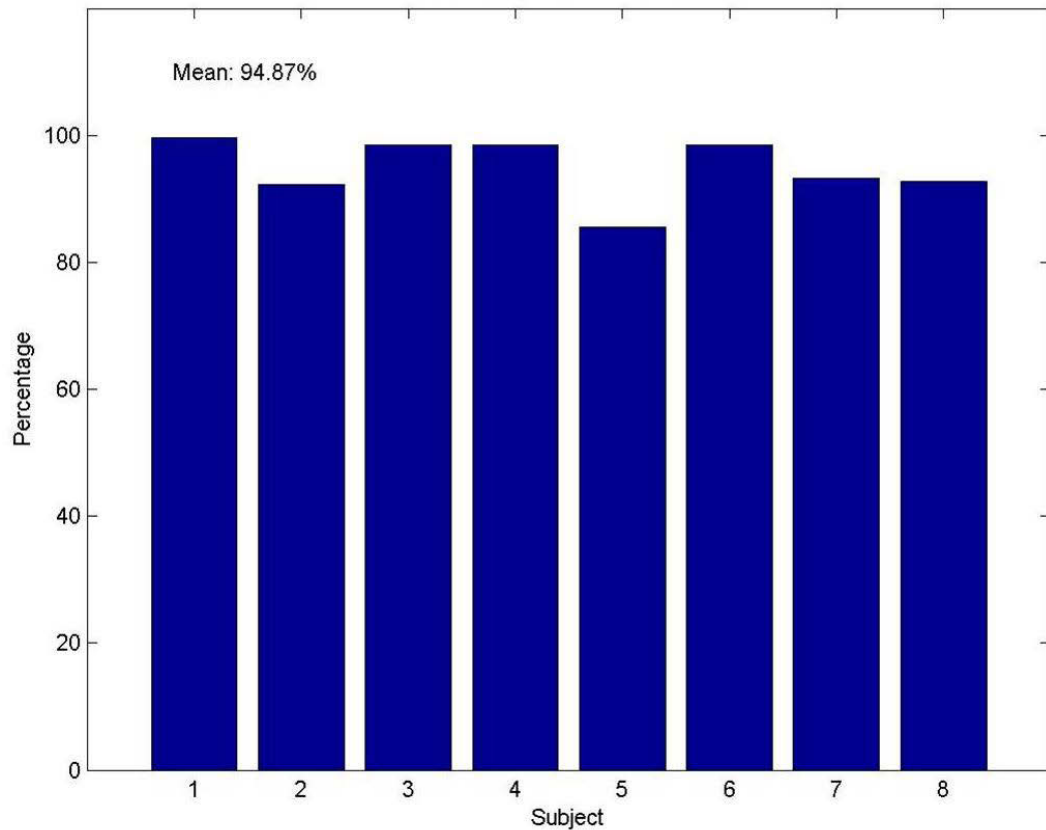


Figure 49 – Percentage of faces visited

The results for proportions of faces visited across 8 subjects. Each bar represents the percentage of faces visited out of all faces and all images for this subject.

As a similar more robust measure of the amount of resources spent looking at faces, we used the number of fixations in face regions. We pooled all the fixations within each image and counted how many of them fell inside a face region of interest. We used the total number of fixation in each image, which varies between 2 fixations up to 12 fixations as 100%, and measured the percentage of fixations that were landing in the face. See Figure 50.

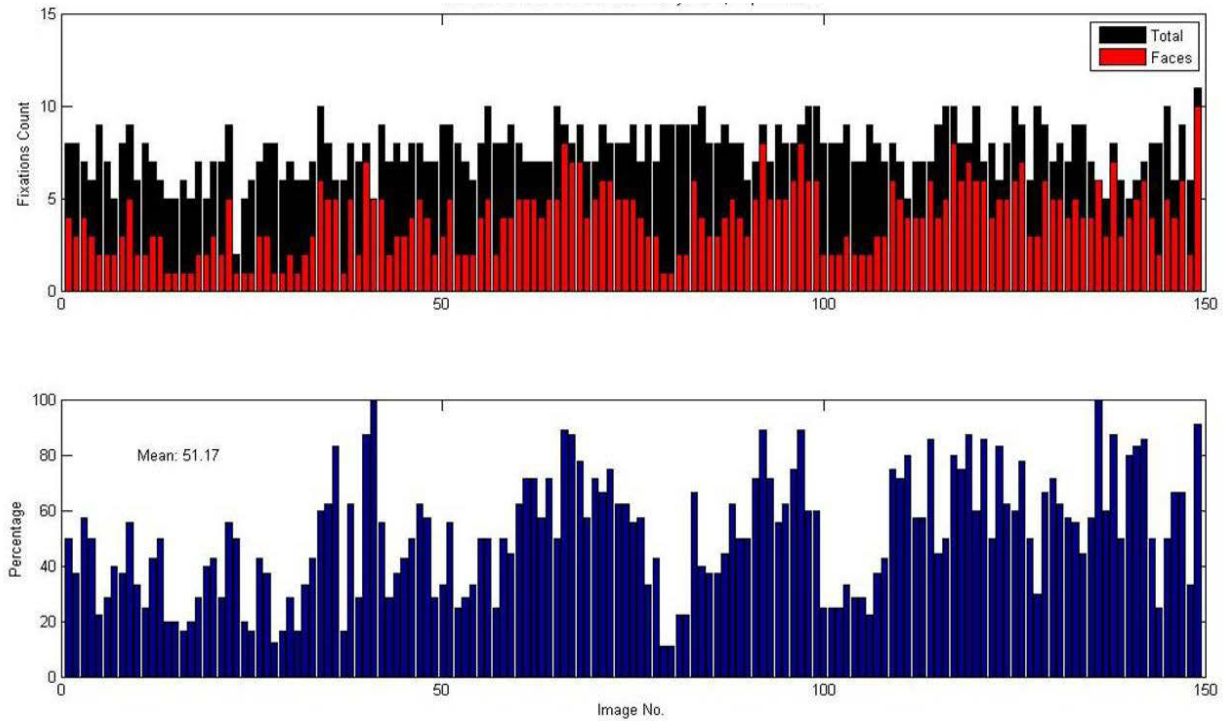


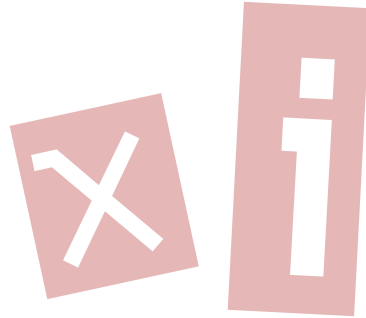
Figure 50 - Number of fixation in faces for subject 1, experiment 1

Top panel. Total number of fixations in each image out of 150 images with faces for subject 1 (chosen arbitrarily). The number of fixations varies. Average number of fixations is 7. Black bars show the number of fixations in each image, and the red bar shows the number of fixations out of these that landed on the face region of interest.

Bottom panel. The proportion of fixations in each image that landed inside a face (red bar divided by blue bar in top panel)

Discussion

First, we demonstrated that in natural scenes containing frontal shots of people, faces were fixated on within the first few fixations, whether subjects had to grade an image on interest value or search it for a specific, possibly nonface target. This powerful trend motivates the introduction of a new saliency model, which combined the “bottom-up” feature channels of color, orientation, and intensity, with a special face-detection channel, based on the Viola & Jones algorithm.



Attention and High-Level Cues.

Part II – Text

*The first forty years of life give us
the text; the next thirty supply the
commentary on it*

Arthur Schopenhauer

H¹⁰uman visual attention serves to delegate the resources of the brain to quickly and efficiently process the vast amount of information that is available in the environment (James, 1890). This process is one of the best modeled functions of the brain; however the model's predictive powers have not reached their full potential (Peters et al., 2005). One of the dominant sensory-driven models of attention currently focuses on low-level attributes of the visual scene to evaluate the most salient areas. Features such as luminance, orientation, and color are commonly combined to make maps through center-surround filtering at multi-scaled resolutions. These maps are normalized to create an overall saliency map, which predicts human fixations significantly above chance (Itti & Koch, 2000; Parkhurst et al., 2002; Peters et al., 2005; Tatler et al., 2005). Filling some of the gap between the saliency map model's current predictive power and the theoretical optimum is thought to be possible by incorporating higher-order statistics in saliency models (W. R. Einhauser, U., Frady, E.P., Nadler, S., Konig, P., Koch, C., 2006). One way of doing this is by adding new feature channels that represent high-level attractive stimuli. For example, recent work has shown that faces, a high-level stimulus, are very attractive (Cerf, Harel, Einhauser, & Koch, 2008). This makes sense, as faces were an important feature to pick out in our evolutionary environment, and thus a process to allocate attention toward faces has reason to arise in the brain.

¹⁰ This chapter is partially based on: Cerf M, Frady P, Koch C, "Faces and text attract gaze independent of the task: Experimental data and computer model", *Journal of Vision*, 2009

Introduction

This study looks at how text, another high-level stimulus, effects human attention in still images and analyzes some of the problems that arise when trying to study text and attention. Attention is measured by examining the locations fixations of subjects using an infrared eye-tracker. This is possible because in normal circumstances a person's eye-position is tightly correlated to the object that the subject is attending (Rizzolatti et al., 1987). This process of measuring attention, though, presents several interesting issues and questions when analyzing a stimulus such as text. One question is reading – reading text is a process that may manipulate attention, since reading requires fixations. Text also conveys information and another question is whether or not meaning could have an effect on attention. To address these issues, two categories of images were created. Images with text objects were taken from the internet and the text was matched and replaced with random letters drawn independently from each of their appearance frequencies in English (e.g., the letter “e” had the highest probability of being drawn). Nine subjects were shown the images with scrambled text and nine subjects were shown images with normal text.

Methods

Subjects

Eighteen subjects were taken from the Caltech community. Subject's ages ranged from 18 to 23, and all subjects were literate. All subjects had either normal or corrected-to-normal vision, were naïve to the purpose of the experiment, and were compensated for participation.

Stimuli

Forty Images were taken from the internet. All images had some form of legible text in them. For the experimental images, the text of the images was removed and replaced with an editable string that was matched as closely as possible for size, font, color, orientation, and shape. Two classes of images were formed. The first class, or the “Normal” class, was created by taking the editable matched text and using the exact same letters as in the original images – these images were to look as if they had not been manipulated at all. The second class, or the “Scrambled” class, was created by replacing each letter independently with a random letter drawn from an English-based distribution of letters. Occasionally, replacements were made to make sure the text looked natural – for example w's and m's take up a larger amount of space than other letters and l's and i's take up a smaller amount of space, if many of these types of letters were in a single word they might have to be replaced or the string recreated so that the word would take up the same amount of space. The images the subjects saw had the exact same font, color, and letter

spacing – since both types were created from a font (subjects never saw the original images).

Experimental paradigm

The experiment consisted of an instruction screen, a calibration period, and the display of images. After the display of each image, the subjects were asked to rate, on a scale from 1-9, how interesting they found the previous image. Each image was shown for approximately 2 seconds in a random order. A fixation cross was shown to each subject which would wait for the subject's fixation. This forced all subjects to start in the center and allowed a check for drift. If the eye tracker did not calculate the subjects eye position to be on the target for any trial, then the subject could be recalibrated. All stimuli were computed and displayed using Matlab (MathWorks, Natick, MA) and its psychophysics toolbox extension (Brainard, 1996). These stimuli were presented on a 19' Dell monitor, whose maximum luminance was 29 cd/m^2 , located 85 cm in front of the subject. Subjects' eye positions were recorded using a noninvasive infrared eye tracker – the Eyelink-1000 system. The manufacturer's software was used for calibration and validation of the fixations.



Figure 51 – Examples of an image transformed used in this experiment

The image on the left contains the normal text matched in size, color, orientation and font. The image on the right has the same attributes as the other, only the letters have been replaced by letters drawn from a random distribution.



Figure 52 – Images with subject's scan paths overlaid

The image on the left contains the scan path of a single subject, and the image on the right shows that of the 9 subjects that saw the normal version of this picture.

Analysis was done on only valid fixations, as several kinds of fixations were either errors or part of the experimental paradigm – e.g., the first fixation was typically removed from the trial. Invalid fixations were defined by a few measures: a first fixation located in the center of the image, a fixation that started before the image was showing, and a fixation that did not fall within the boundaries of the image. On very rare instances was a subject's first real fixation considered valid – it must start after the display of the image and not be located in the center. This is possible through the subject moving his eyes immediately after the fixation cross, but it happened very few times.

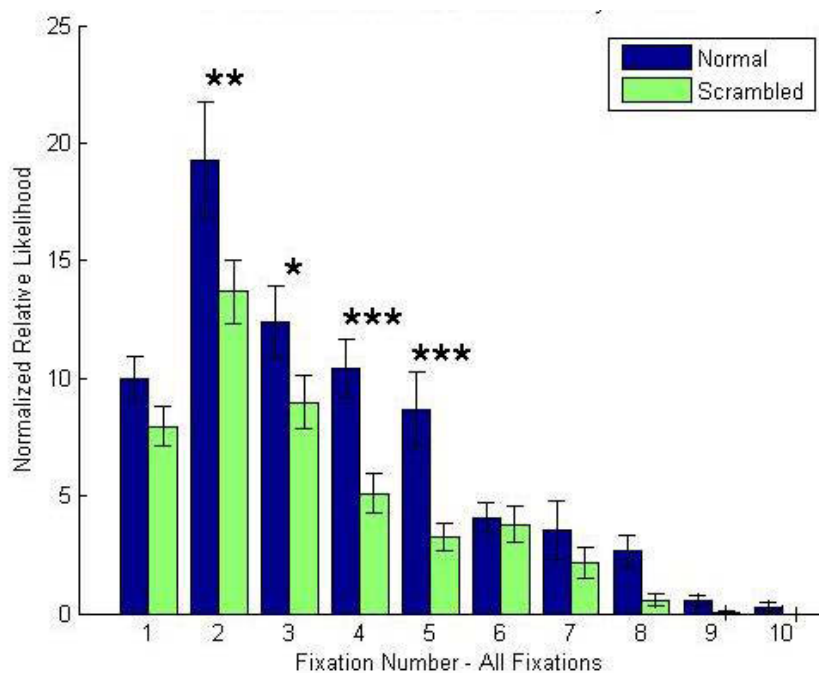
Results

Areas called regions of interest were created around each text object, which served to report that a subject was looking at a text object when his fixation landed inside this region of interest. The ROI also allowed us to calculate a chance value for a valid fixation (see methods for definition of valid fixation) to fall on this certain text object. The chance value is the probability that a certain text object would be fixated on by chance, and is based on a nonbiased distribution of fixations. This probability is calculated for a specific image's ROI by taking all the fixations from all other images and calculating what fraction fall in the region of interest for that particular image (this type of normalization accounts for both the text object's size and position). Thus we can compare the fraction of fixations on a text object with the value of a fixation hitting that region by chance. The experimental results show that both normal text and scrambled text are much more likely to be fixated on than chance ($p < 10^{-6}$).

This chance value can also be a normalization factor allowing us to compare ROIs that differ in size and position (as these factors both influence how likely a certain region is to be fixated by chance). By dividing the fraction of fixations in each text object's ROI by its chance value, we form a normalized measurement. Under certain assumptions – e.g., fixations are independent – this value is analogous to the number of times more likely the ROI is attended than chance. Our results show that fixations falling on normal text produce a normalized value of 11.2 and scrambled text produces a value of about 7.5. These two values are significantly different from each other ($p < 0.02$). A possible explanation to this discrepancy is fixations due to reading. I defined “reading fixations” as a fixation in the ROI given that the previous fixation was in the same ROI. Removing reading fixations from the scan paths removes all significant differences

measured from the two types of viewers, and reduces the normalized value to 7.2 for normal text and to 5.5 for scrambled text ($p > 0.25$). These fixations were removed from the data set and consisted of approximately 11 percent of all fixations, and 40 percent of fixations that were in the region of interest.

This kind of reading effect is also apparent when looking at the data fixation by fixation. Examining all fixations the fraction of trials in which the first, second, third, etc., fixations are in the region of interest shows that the main differences from normal and scrambled text come from the second through fifth fixations. If one examines only the fraction of trials that have fixations in the ROI for the first time the data shows that there is no significant difference throughout the fixation numbers (See Figure 53). These results again suggest that as far as attracting attention normal and scrambled text are similar, but many more extra successive fixations are invoked by normal text than by scrambled text.



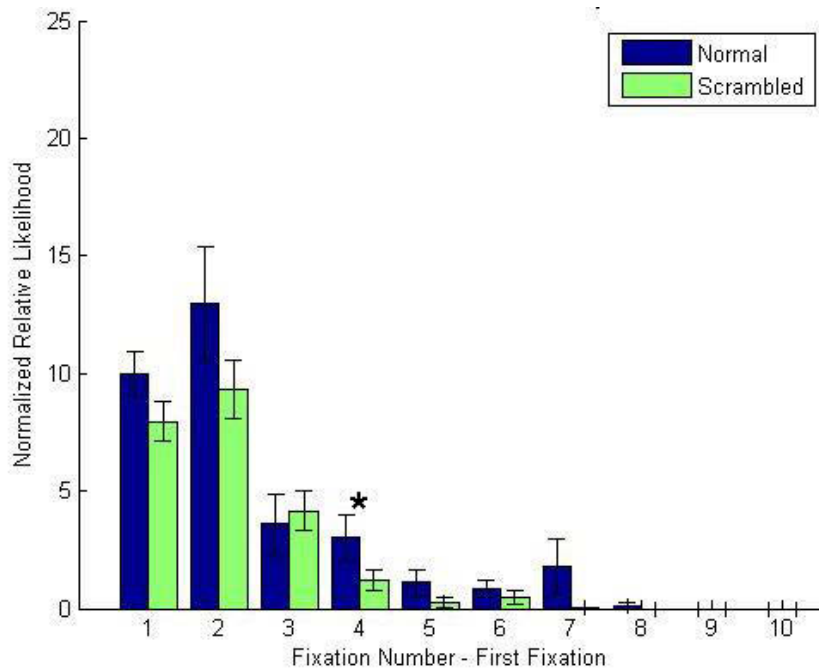


Figure 53 – Comparison of fraction of trials with fixations in the region of interest

Upper panel. All fixations

Lower panel. Only the first fixations in the ROI. (Significance Key: * = $p < 0.10$, ** = $p < 0.05$, *** = $p < 0.01$). These graphs show that the main difference between Normal and Scrambled text lies in successive fixation to the ROI in the second through fifth fixation.

To analyze if text contributes more than just its low-level attributes to power attention we compare how well the standard low-level feature-driven saliency map does in comparison with this map plus an extra high-level text channel. Our standard saliency map consists of three channels: orientation, luminance, and color.

$$s = \frac{1}{3}(O + L + C)$$

The modified map is made with the extra text channel.

$$s = \frac{1}{4}(O + L + C + T)$$

The relative weight between the text channel and the other 3 is actually a free parameter, however the performance does not significantly change with a higher relative weight – what is actually important is mainly the addition of this information to the model (see Figure 55). The text channel was simply the ROI maps created earlier, it is a binary image where 1 represents text at this or near this location and 0 represents no text. The text channel will in the future be implemented by a text detector, but this analysis has followed an ideal text detector – one implemented by a human. Performance of the saliency map was measured by its ability to predict fixations using the receiver operating characteristic (ROC). The hit rate is calculated by determining the locations at which the saliency map is above threshold and if there are fixations at these locations, similarly the false alarm rate was calculated by measuring the locations at which the saliency map was above threshold and there were no fixations. The ROC curve was created by varying the threshold to cover all possible ranges of values the saliency map produces. The area under the curve (AUC) is the general measure of how well the saliency map predicted fixations, and an area of 50% reflects chance and an area of 100% reflects perfect predictions.

The performance of the standard saliency map was on average 72.1% AUC for the 40 images. Adding the text channel improved this value to 77.3% AUC; this is a significant improvement

($p = 0.0038$, one-tailed t-test). The text channel leads to improvements for every image. Predictions of fixation location for subjects viewing normal text improved from 71.4% AUC to 77.5% AUC, and scrambled type improved from 72.9 to 77.1. Both results are significant improvements ($p = 0.0022$, 0.0113 , respectively). Removing reading fixations from the scan paths still leads to significant improvements for both types of subjects. Predictions for subjects viewing normal text improves from 70.3% to 74.3% ($p = 0.0275$), and for scrambled text improves from 72.2% to 75.4% ($p = 0.0413$). Figure 54 shows clearly that adding a text channel to the model will improve performance.

The results from the saliency map suggest that text is in actuality more salient than predicted by its low-level features. This experiment, however, does not definitively prove that there is a text channel in your brain, or even that your brain learns to tune into the texture of words. A possible way to prove this is by doing similar experiments on subjects who have never been exposed to text objects (while also normalizing for age and intelligence), and observe a much lower frequency of fixations to text regions. The experiment does suggest that text is overly salient, beyond the saliency of its low-level components, and a learning response for bottom-up attention in the brain is likely.

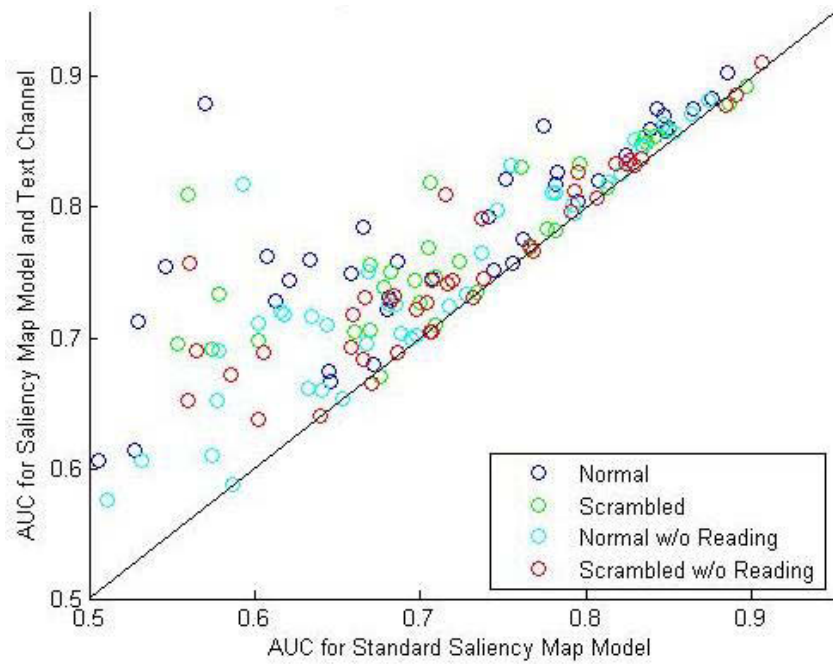


Figure 54 – Comparison of model performance

Circles above the black dividing line indicate an improvement in the Saliency Map Model when adding a text channel. Each circle represents the model's performance of predicting the 9 subjects' fixations on a particular image. Performance improved for all images when subjects saw normal text with and without reading fixations, 90% of images with scrambled text, and 80% of images with scrambled text and without reading fixations.

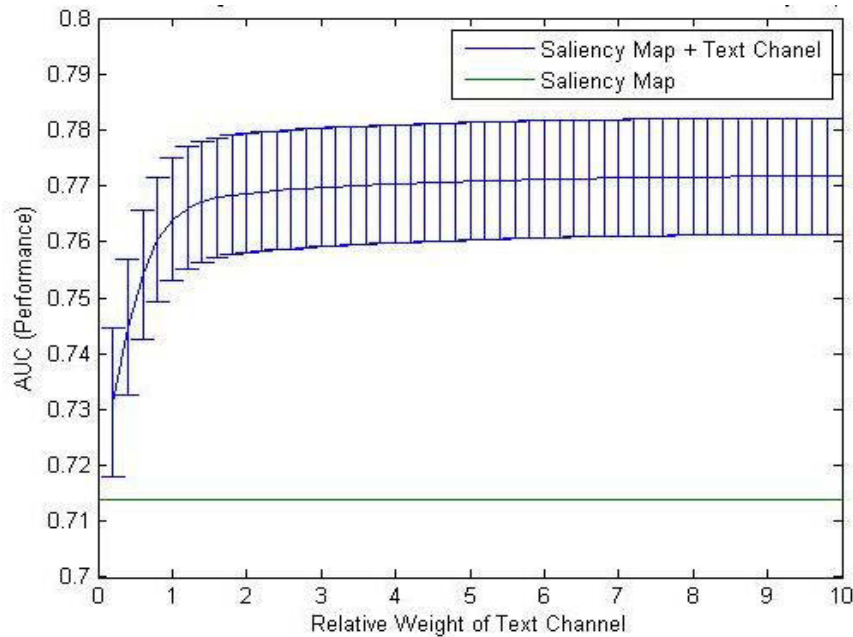


Figure 55 – Relative weight of text channel versus

Adding more weight to the text channel does not significantly improve the model above a relative weight of 1

There are many other interesting questions to answer when dealing with text and other high-order stimuli. The next steps in this research we are currently investigating are to compare the statistics from these experiments with faces and with a nonspecial object – we decided a cell phone. Hopefully, we will find many similarities between faces and text (both of which we believe are special), and many differences with the cell phone. We are also currently going to study whether faces, text, and cell phones are avoidable – e.g., you are asked to search for a target and you know that it cannot appear on a face. Our hypothesis is that since faces and text are such strong stimuli, you will not be able to ignore them in your search, but that you will be able to ignore cell phones.

The saliency map model used was taken from the Saliency Toolbox (D. Walther, Koch, C., 2006). The program takes a color image and produces three maps with values in a range of 0 to 1, where each map represents a certain low-level statistic thought to be important to saliency: luminance contrast, orientations, and color contrast. Higher values represent locations that are calculated to have more saliency. Each of the 40 images had a binary map created that served to report locations at which there was text. This map was made by forming a box around the text objects added to the image. The box was filled to cover all parts of the text and typically extended 15 pixels around the edges of each text object. The maps both served as the region of interest maps – used to quantify the fixations that were targeted at text objects – and as the text channel to add to the saliency map. The text channel was linearly added to the saliency map with an equal weight to produce the advanced model.

The performance of the saliency map was calculated using a receiver operator characteristic curve, which tries to establish an unbiased representation of a binary response. The hit rate is calculated by finding the locations at which the saliency map is above threshold and there is a fixation and dividing this number by the total number of fixations. The false alarm rate is similarly calculated by finding locations at which the saliency map is above threshold and there is not a fixation and dividing this number by the total number of locations. These values are plotted against each other as the threshold is varied to cover all levels of the saliency map's output. The area under the curve is a general measure for the overall performance of the map, and is calculated by taking the integral of the roc curve.



Attention and High-Level Cues.

Part III – Faces and Text

Her eyes were like two brown circles with big black dots in the center.

Russell Beland, Springfield

*(from the ETNI - English Teachers Network Institute -
"Worst analogies ever written in a high school essay"
<http://www.etni.org.il/farside/analogies.htm>)*

V¹¹isual attention serves to delegate the resources of the brain to quickly and efficiently process the vast amount of information that is available in the environment (James, 1890). Certain aspects of selective visual attention, in particular task-independent, exogenous and bottom-up driven attention, has become reasonably well understood and quantitative models have been derived to explain attentional and eye movement deployments in the visual scene (L. Itti & C. Koch, 2001). In general, it is thought that observers' fixational eye patterns correlate tightly with their covert attention under natural viewing conditions (Einhäuser et al., 2006; Rizzolatti et al., 1987).

Commonalities between different individuals' fixation patterns allow computational models to predict where people look, and the order in which they view different items (Cerf, Cleary, Peters, Einhäuser, & Koch, 2007; Foulsham & Underwood, 2008; A. Oliva, Torralba, Castelano, & Henderson, 2003). There are several models for predicting observers' fixations inspired by putative neural mechanisms (Dickinson et al., 1994). One of the dominant sensory-driven models of attention focuses on low-level attributes of the visual scene to evaluate the most salient areas. Features such as intensity, orientation, and color are combined to produce maps through center-surround filtering at multi-scaled resolutions. These maps of feature contrast are normalized and combined to create an overall saliency map, which predicts human fixations significantly above chance (Itti & Koch, 2000). Filling some of the gaps between the

¹¹ This chapter is partially based on: Cerf M, Frady P, Koch C, "Faces and text attract gaze independent of the task: Experimental data and computer model", *Journal of Vision*, 9(12):10, 1-15, 2009

predictive power of current saliency map models and their theoretical optimum is the incorporation of higher-order statistics (Einhäuser et al., 2006). One way of doing this is by adding new feature channels for faces or text into the saliency map.

Visual attention is frequently deployed to faces, to the detriment of other visual stimuli (Bindemann, Burton, Hooge, Jenkins, & de Haan, 2005; Bindemann, Burton, Langton, Schweinberger, & Doherty, 2007; Cerf, Harel, Einhauser et al., 2008; Mack, Pappas, Silverman, & Gay, 2002; Ro, Russell, & Lavie, 2001; Theeuwes & Van der Stigchel, 2006; Vuilleumier, 2000). Evidence from infants as young as 6 weeks old suggests that faces are visually captivating (Cashon & Cohen, 2003). We here investigate the attractiveness of faces in the context of detecting a target in natural scenes where faces are embedded in the images.

Text is yet another entity that frequently captures humans' gaze in natural scenes (Cerf, Frady, & Koch, 2008). Although the claim that one is born with an innate ability to detect faces is still under debate (Golarai et al., 2007; Simion & Shimojo, 2006), it is unlikely that we are born with a similar mechanism for text or cell phones detection. Any skill in detecting these man-made items in a natural scene should be attributable to experience.

This study considers how faces, text and complex man-made objects (cell phones) affect gaze in still images. We track subjects' eye movements in a free viewing task and in multiple search tasks to measure the extent to which subjects can avoid looking at faces and other objects. Such

tasks are close to day-to-day situations and shed light on the way attention is allocated to important semantic entities.

Methods

Experimental procedures

Subjects viewed a set of images (1024 x 768 pixels) in four experiments. The general structure of all tasks is the same. Prior to each session, the subjects' gaze was determined through a calibration process. Before each stimulus onset, the subjects were instructed to look at a white cross at the center of a gray screen. If the calculated gaze position was not at the center of the screen, the calibration process was repeated to ensure that position was consistent throughout the experiment. Eye-position data were acquired at 1000 Hz using an Eyelink 1000 (SR Research, Osgoode, Canada) eye-tracking device. The images were presented on a CRT2 screen (120 Hz), using Matlab's Psychophysics and eyelink toolbox extensions (Brainard, 1996; Cornelissen et al., 2002). Stimulus luminance was linear in pixel values. The distance between the screen and the subject was 80 cm, giving a total visual angle for each image of 28°x21°. Subjects used a chin-rest to stabilize their head. Eye movement data were acquired from the right eye alone. All subjects had normal or corrected-to-normal eyesight. All subjects were naïve to the purpose of the experiment. These experiments were undertaken with the understanding and written consent of each subject. All experimental procedures were approved by Caltech's Institutional Review Board.

In the first ("free-viewing") experiment, 27 subjects viewed 1 out of 3 categories of images, with 9 subjects per category. The categories were natural scenes that contained one or more faces (157 images), one or more discrete text elements (37 images), or one cell phone (37 images).

Images were presented to the subjects for 2 seconds, after which they were instructed to answer the question, “How interesting was the previous image?” using a scale of 1-9 (9 being the most interesting). Subjects were not instructed to look at anything in particular; their only task was to rate the entire image. Image order within each block was randomized throughout the experiment. Figure 56 shows a sample image from each of the 3 categories.

In order to provide a fully equivalent image group, for comparison between the categories independent of size and background, we had six additional subjects perform experiment 2 (“control for the relative effect of size”) where a set of 25 images was presented in a “free viewing” task. These images had faces, text or cell phones artificially embedded such that they occupied the exact same size and location in the image.

In a third (“search”) experiment, 15 additional subjects were instructed to look for a target – two concentric circles (50x50 pixels; 1.36°x1.36° visual angle) surrounding a cross-hair (Figure 56b) – embedded in the image. The cross's location was selected randomly from a uniform distribution. Each block had 74 images from a single category, 37 of the images were the same as in the “free-viewing” experiment, and the other 37 were identical images except that they included the cross embedded somewhere within the image. Images were presented for 2 seconds after which subjects were given a two-alternative-forced-choice (2AFC) question asking “Was the cross in the previous image (y/n)”. Seven of the 15 subjects performed under a “free search” instruction, in which they were told that “The target can be located anywhere in the following images”. The remaining 8 subjects performed the search task under an “avoid” instruction, in which they were told, “The target cannot be located on face objects in the

following images” for the face block, “text objects” for the text block, and “cell phone objects” for the cell phone block. The target was placed outside of these regions accordingly. Before the experiment, subjects were given 12 training trials using separate images. Subjects were told the location of the cross at the end of each practice trial in order to familiarize them with the looks of the cross-hair and difficulty level.

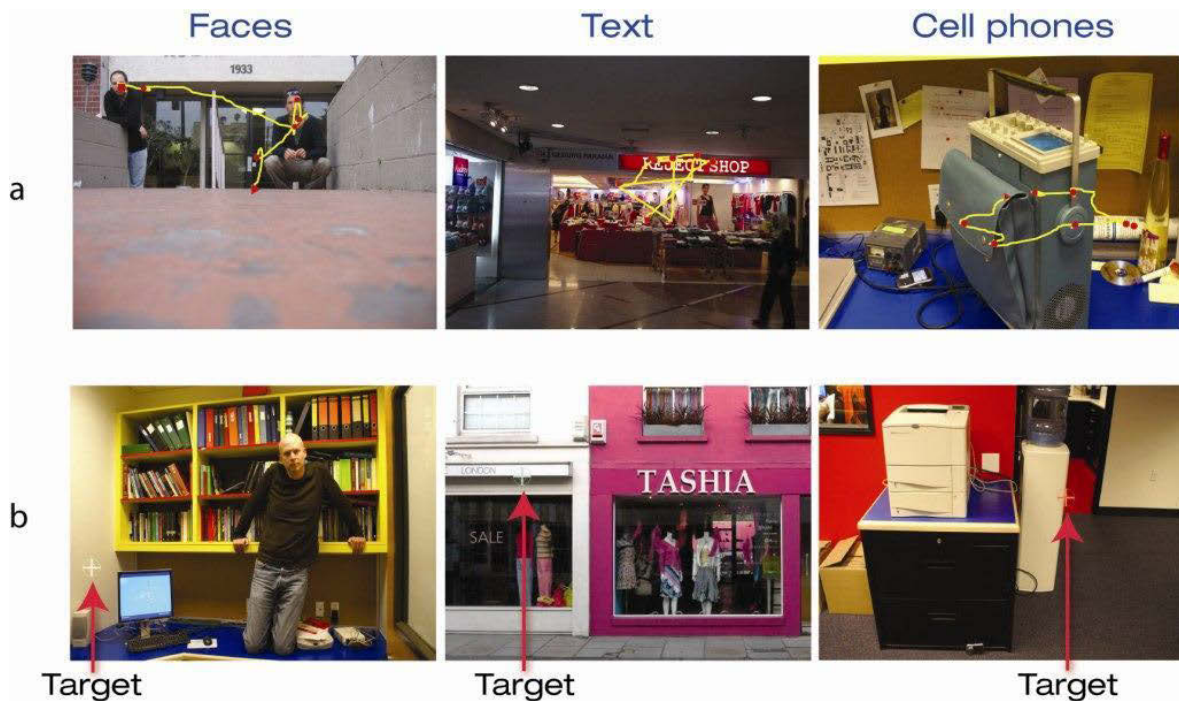


Figure 56 - Examples of images from the three categories: Faces, text, and cell phones

a. Examples of images from the three categories: Faces, Text and Cell phones, with scanpaths of one individual from each of the 3 groups superimposed. The triangle marks the first and the square the last fixation, the yellow line the scanpath, and the red circles the subsequent fixations. The trend of visiting the faces and text first - typically within the 1st or 2nd fixation - is evident, while cell phones do not draw eye gaze in the same manner.

b. Examples of images used in experiment 3 (“search”). Red arrows point to the superimposed target cross.

Finally, four additional subjects performed a search task in a 4th experiment (“control for the effect of adaptation”) where images from each category were intermixed, creating one long experiment. The experiments included both the free search task and avoid task instructions such that prior to each trial the instructions could be either “The target will not appear on a face”, “on a text”, or “on a cell phone object” or “The target can appear anywhere”. This experiment controlled for a general adaptation of strategy that may occur over a block of the same image types. Furthermore, by looking only at the free search trials in this experiment, we can rule out any possible top-down influences as the subjects are unaware of the coming stimulus category prior to the image presentation.

Images

All stimuli were designed or chosen as images that are representative of a real-world scene.

The face images were manually photographed in indoor and outdoor environments. Images included people (and their faces) of various skin colors, age, and postural positions. A few images had face-like objects (Smiley t-shirt, animal faces, and objects that had irregular faces in them - masks or the Egyptian Sphinx). The text images were taken from the internet and were chosen such that only a few text objects appeared in the scene. Images containing a cell phone were manually photographed in an office setting where a cell phone would be considered a reasonable item. Cell phones were chosen to represent an object that is less important to a

human's visual environment. The majority of the images contained only one of the three entities.

The average face size was $5\% \pm 4\%$ ($5.42^\circ \pm 4.84^\circ$ visual angle); the average size of text was $6\% \pm 1\%$ ($5.93^\circ \pm 2.42^\circ$ visual angle); and the average size of cell phones was $1\% \pm 0.4\%$ of the entire image ($2.42^\circ \pm 1.53^\circ$ visual angle).

In the search tasks, the cross's color was varied such that it would be challenging for the viewer to find, but obvious to recognize once it was located. The color was altered throughout the experiment such that observers could not solve the task by simply looking for a specific color. After choosing a random location for the cross (given the constraints of each block), the cross's color was determined by taking the average of all pixel colors in a local (50x50 pixels) neighborhood and then increasing the brightness by 10% (this made the cross visible in all locations in which it was placed, while still making it challenging to find). These parameters were chosen to yield about 70% success rate on 2AFC test trials.

All the fixations, scanpaths and images used in the experiment are available online at <http://www.fifadb.com> for further studies (see appendix for information on usage).

Analysis Metrics

Fixations

Fixations were determined by the built-in software of our eye-tracking system. We categorized fixations by their “fixation number” based on a fixation’s position in the ordered sequence of fixations (*i.e.* first, second, third). The “initial fixation” is the fixation occurring before stimulus onset, when the subjects are focusing on the centered fixation cross, and is not counted as part of the ordered sequence of fixations.

We calculate the fraction of fixations in an ROI by summing the number of fixations that fall within this region over all trials and then dividing by the number of trials. A trial consists of one subject being presented one image, and will contain one ordered sequence of fixations.

Saccades

Saccades were also determined by the eye-tracking system. The “saccade planning time” is the duration of time between the stimulus onset and the initiation of the first saccade. Saccade planning times smaller than 50 ms or greater than 600 ms were discarded to remove outliers and artifacts. The duration of viewing time was measured based on when the saccade started as opposed to when the next fixation started so that timing would not be affected by the length of the saccade, or the distance to the target.

Saccades were categorized as being “on-target” or “off-target” depending on the location of the following fixation. If the fixation fell in an ROI (*i.e.* on a face) then the saccade was considered to be on-target. If the fixation fell outside a region of interest, then the saccade was categorized as off-target. See Figure 57 for illustration.

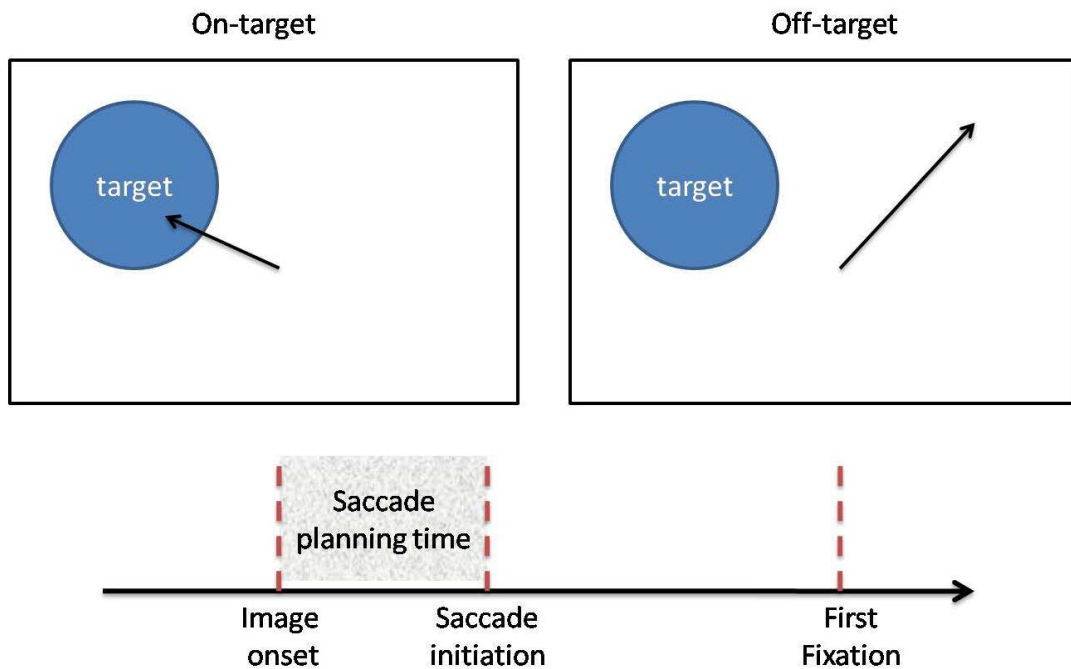


Figure 57 - Illustration of the computation of the saccade planning time

For each subject and image, we take the time spent before the initiation of the first fixation as the “saccade planning time”. The location of the saccade’s ending point is used to bin the times into two groups based on whether or not the saccade landed in a region of interest. If the first fixation was on a face / text / cell phone we define it as on-target, and if it is anywhere else we define it as off-target.

Baseline calculation

To compute chance level of performance, we calculated the ratio of the fraction of fixations that land in a target's region of interest to the fraction of a baseline distribution. The baseline for a particular image is the fraction of all subjects' fixations from all other images that fall in the ROI of the particular image (see Figure 58 for illustration of the baseline calculation). This takes into account the varying size and locations of the ROI in all images (as these factors both influence how likely a certain region is to be fixated on by chance), and the double spatial bias of photographer and observer (Tatler et al., 2005). As the baseline value estimates the chance level of landing in a given ROI, we make a comparison with this baseline value to the data by dividing the fraction of fixations landing in the ROI by its baseline value.

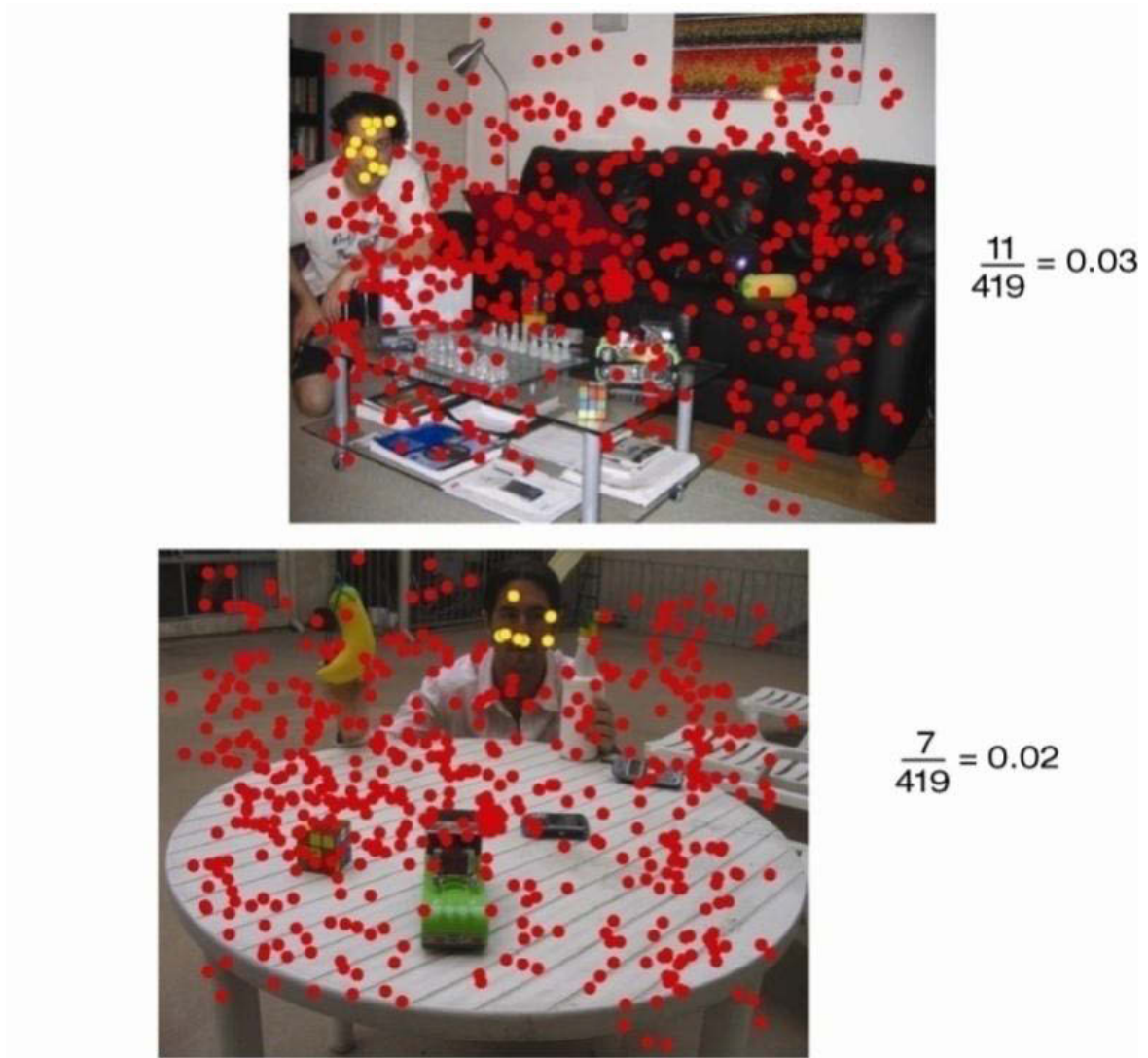


Figure 58 – Illustration of the computation of the baseline

For each subject we consider all fixations, except the ones recorded for this image. We then compute the fraction of these which fall within the ROI for this image. Here we see the baseline calculation for two images containing faces. For each image, the fixations from all other images for that subject are superimposed in red and yellow. Yellow dots indicate that the fixation falls within the ROI. For the bottom image, seven fixations out of 419 fell within the

ROI. The average number of fixations in the ROI for all face images using this baseline calculation was 0.4%.

Results

Psychophysical results

Experiment 1 (“free-viewing”)

To evaluate the results of the 27 subjects' free-viewing of the images, we looked at both the fixations and the saccade planning times of each subject. The locations of the fixations were compared with minimally sized rectangular ROIs manually defined around each target object – face, text, and cell phone in each image in the entire collection.

During free-viewing, subjects fixated on faces within the first two fixations in 89.3% of the trials, which is significantly above chance ($p < 10^{-5}$, t-test, Figure 59). Similar measures for the text show that subjects fixated on text within the first two fixations in 65.1% of the trials ($p < 10^{-5}$, t-test). Overall, faces and text are much more likely to be fixated within the first two fixations than is predicted by chance ($p < 10^{-10}$, t-test). In contrast, by the 2nd fixation, cell phones were visited in only 8.4% of the trials. The on-target saccade planning time for face images was very rapid (203 ± 57 ms). Off-target saccade planning time was equally rapid (199 ± 72 ms). On-target saccade planning time for text was not as rapid as for faces, but significantly faster than for cell phones (239 ± 54 ms for text; 313 ± 47 ms for cell phones; $p < 10^{-7}$ for faces and text compared to cell phones and faces compared to text, t-test). Off-target saccade planning time was equally rapid for text images (241 ± 77 ms). Off-target saccade planning time for cell phones was 290 ± 89 ms. Given that the ROIs were chosen very conservatively (*i.e.* fixations

just next to a face did not count as fixations on the face), these results show that faces and text are highly attractive.

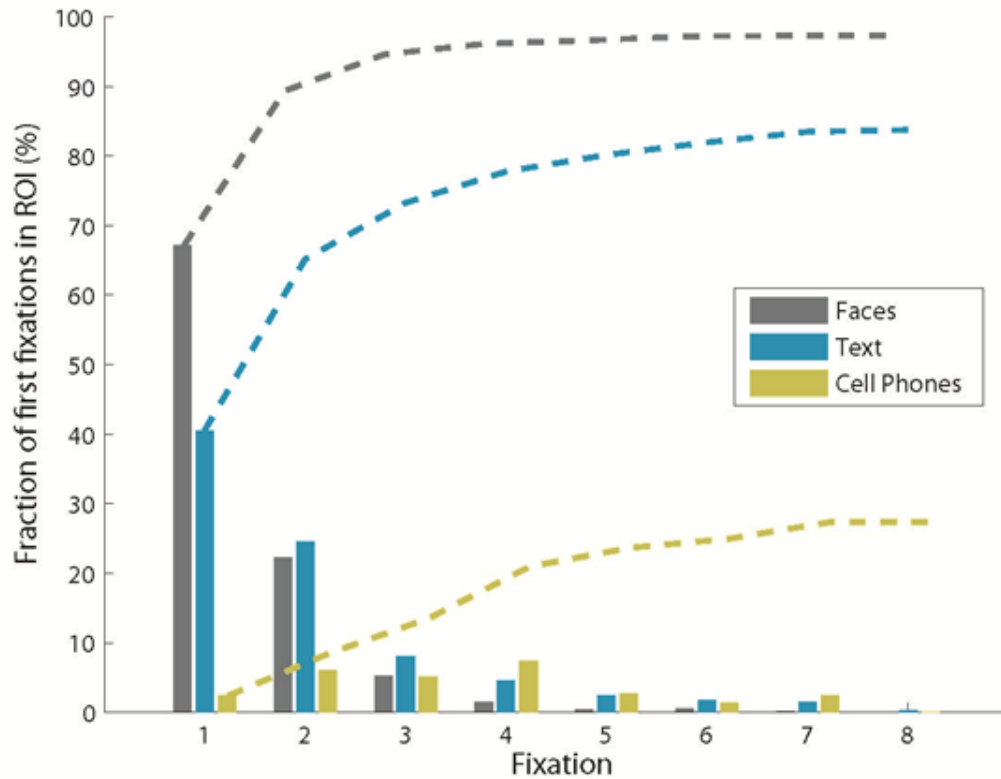


Figure 59 - Extent of first fixation on ROI during free-viewing task

Bars depict percentage of trials which reach the ROI the first time in the first, second, third... fixation. The dashed curves depict the integral, *i.e.* the fraction of trials in which faces were fixated on at least once up to and including the n^{th} fixation. This data shows that subjects tend to fixate on faces much earlier than text, and that they fixate on both text and faces earlier than cell phones.

Experiment 2 (“control for the relative effect of size”)

To be certain that the attractiveness of faces, text, and cell phones is not due to size, position, or competing stimuli in the background, we ran a second, follow-up experiment controlling for each of these factors. Six new subjects were shown a dataset comprised of the three entities artificially embedded in the same images (background-wise), such that the entities occupied the same size and position. Subjects engaging in free-viewing of these images repeat the same tendency to look at faces the most (61% of the first fixation went to faces), more than text (44% of first fixation landed on a text region), and much more than cell phones (7%). Comparing these results to chance performance, calculated using our baseline distribution of fixations, we see that faces are 16.6 times more likely to be fixated on than the baseline, while text is 11.1 times more likely, and cell phones are only 3.2. There is a significant difference between faces and text ($p < 10^{-6}$, t-test), faces and cell phones ($p < 10^{-10}$, t-test), and text and cell phones ($p < 10^{-8}$, t-test). This strengthens the claim that faces are indeed more attractive to viewers than text, which in turn is more attractive than cell phones. The large difference between the three entities shows that the bias towards faces is due to the higher attractiveness of these, as the shared background among all entities makes the comparison controlled.

We observe similar trends in the saccade planning times for this experiment for faces (220 ± 51 ms for on-target; 208 ± 66 ms for off-target), for text (233 ± 59 ms, on-target; 236 ± 65 ms, off-target), and for cell phones (299 ± 53 ms, on-target; 310 ± 42 ms, off-target). Again we see that there is a global depression in saccade planning time when there is a face in the image, regardless of whether the saccade is on- or off-target.

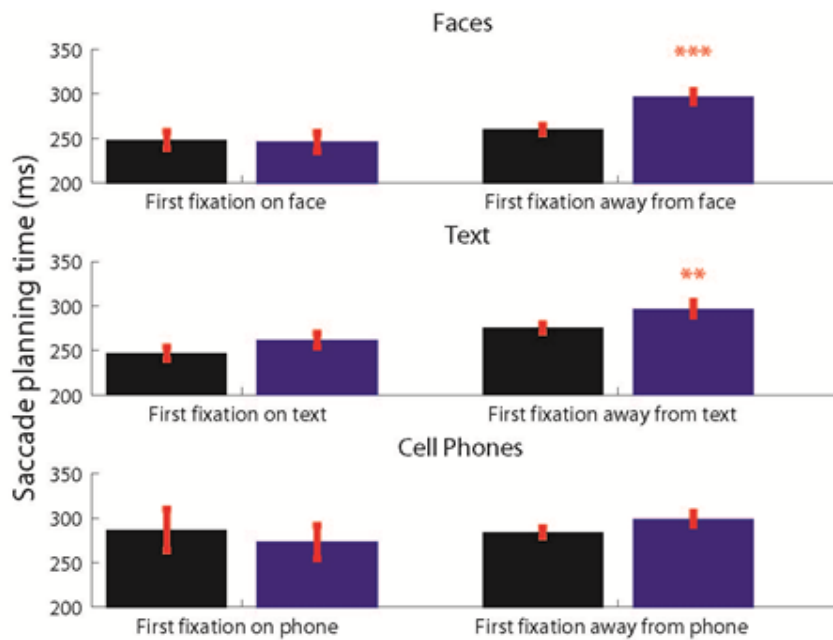
Experiment 3 (“search”)

To assess what affect the two different search tasks (“free search” and “avoid”) had on subjects’ allocation of gaze, we compared fixational patterns between the two tasks. In the free search task, faces were fixated within the first two fixations 24.0% of the time. When subjects were instructed to avoid looking at a face, faces were nevertheless fixated upon within the first two fixations 27.7% of the time. Text was fixated within the first two fixations 32.1% of the time in the free search task, vs. 37.4% of the time in the avoid task. Cell phones were fixated at approximately the same frequency for both search tasks – 12.2% for free search and 10.1% for avoid search. For both faces and text in both tasks these values are significantly lower than in the free viewing task ($p < 10^{-3}$ for all, *t-test*), but are still higher than cell phones in the free viewing task ($p < 0.001$, for all *t-test*). The fraction of cell phones fixated was not significantly different across the search tasks and the free viewing task.

Furthermore, we compared the timing of the initial fixations (free search and avoid). For the face stimuli, we observed that in the free search task on-target saccade planning times took on average 248 ± 63 ms (Figure 60), while off-target fixations took 260 ± 75 ms. These differences are not significant. In the avoid task, on-target saccade planning times took 246 ± 73 ms on average, but off-target times took 297 ± 93 ms. Subjects in the avoid task take significantly longer to make their first saccade off-target compared to both the avoid task on-target ($p < 10^{-3}$, *t-test*) and the free search task off-target ($p < 10^{-3}$). Simply put, it takes longer to *not* look at a face when told to avoid it, than to *not* look at a face when freely searching the scene. We also observe that when avoidance fails – *i.e.* when subjects look at a face under the avoid instruction

- the time before the saccade initiates is not statistically different as when subjects look at a face under the free search instruction.

Looking at text, we observe a similar pattern in the timing of fixations. Again, the avoid task off-target timing (297 ± 106 ms) is significantly greater than the other 3 conditions - the avoid task on-target (261 ± 61 ms), the free search task off-target (275 ± 77 ms), and the free search task on-target (247 ± 52 ms) ($p < 0.02$ for all). There is no significant difference in the timing when the first fixation goes to a face versus to text. More so, like for faces, there is no significant difference between the timing of the on-target and off-target fixations in the free search task for text. Cell phones timing results show no significant differences between any of the 4 categories of fixations (Table 5).



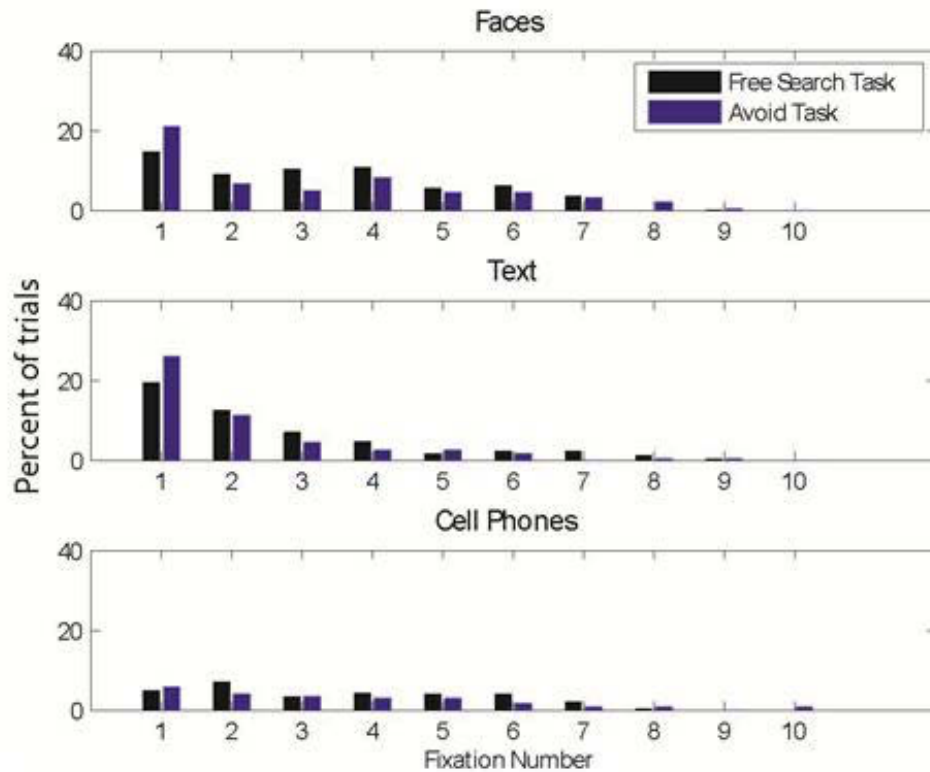


Figure 60 – Saccade planning durations and proportions of fixations in avoid/search task

a. Saccade planning durations across the three stimulus categories. Comparison of duration of first fixation when subjects fixated on the high-level entities (**upper panel** – faces, **middle** – text, **bottom** – cell phones). Left two bars in each panel correspond to subjects' on-target fixations and the right two bars correspond to those that went off-target. Black bars comprise of trials where subjects were instructed that the search target could appear anywhere in the image. Blue bars represent avoid trials. Stars indicate significance for t-test comparison between rightmost bar and the 3 other bars in each panel ($*** = p < 0.001$, $** = p < 0.02$). In the avoid task, fixations away from the face/text (represented by the rightmost bars in each panel) are

significantly higher than each of the other three bars in that panel.

b. Fraction of first on-target fixations in the free search versus avoid search tasks across the 3 stimulus categories. Similar to free-viewing fixations, we show in each panel the fraction of trials in which a fixation landed on the ROI for the first time by the *n*th fixation.

In order to test whether the globally faster first fixations in the face block are due to faces attracting attention faster, rather than a general adoption of a strategy during the ‘face’ block, we further investigated how saccade planning times vary over the course of the experiment. We performed a linear regression relating the trial number to the saccade planning time throughout the course of each block. We found no regression coefficient that was significantly below 0, indicating that none of the entities (faces, text, or cell phones) shows a decrease in saccade planning time for any of the experiments (free view, free search, avoid search).

In line with the overall trend of cell phones having the least significant effect on fixations during search, performance is better for trials where the distracting entity is a cell phone. Subjects were able to correctly identify the presence/absence of the target in $85\% \pm 2\%$ of the trials when the image contained a cell phone, while they could do so in only $82\% \pm 3\%$ of trials containing a face; and $78\% \pm 3\%$ of trials containing text. There is a significant difference in performance between faces and text, as well as between text and cell phones ($p < 10^{-3}$ for both; t-test). In trials where subjects were told to avoid locations where target would not appear, performance decreased in all cases to $72\% \pm 13\%$ for trials containing faces, $72\% \pm 12\%$ for trials containing

text, and $82\% \pm 8\%$ for trials containing cell phones. While the trend of significant difference ($p < 0.01$, t-test) between cell phones and the other two categories remains, the difference between faces and text is eliminated, supporting the claim that whether told to avoid looking at faces or text, subjects employ a similar search mechanism.

Experiment 4 – control for the effect of adaptation

We further tested the absence of change in timing due to adaptation by measuring saccade planning time effects in a conjoint dataset where effects of adaptation should not occur. This dataset was created by stringing together the three types of entities in one long sequence as opposed to three separate blocks. The sequence consisted of both avoid and free search instructions. For the avoid task, we see a consistent trend of the data to that in the original experiment. For saccades going towards faces the saccade planning time was 241 ± 60 ms, for text it was 251 ± 63 , and for cell phones it was 292 ± 73 ms. Off-target timings also show the same increase in planning time for faces (290 ± 69 ms), text (282 ± 73) and even cell phones (310 ± 52 ms) ($p < 0.02$ for faces/text; t-test; cell phones increase in timing is not significant).

We can also use this data to test whether subjects were changing their viewing strategies based on prior knowledge of the stimulus category. The free search set of intermixed trials are preceded by the same instruction regardless of stimulus entity, thus giving no information about the category of the stimulus prior to its presentation. We see that saccade planning time towards faces is 244 ± 69 ms, towards text is 250 ± 99 ms, and towards phones is 299 ± 68 . The

off target timings are also consistent with the results of experiment 3: 260 ± 60 ms for faces, 266 ± 62 for text, and 310 ± 65 for cell phones. This indeed shows that the timing effects are due to the biases these entities pose rather than a general effect of adaptation or category-dependant strategy.

See Table 5 for a list of all performance values and the latency for each entity.

	Faces	Text	Phones	
Saccade Planning Time (on-target)	203 ms	239 ms	313 ms	Free View
	220 ms	233 ms	299 ms	Free View (control)
	248 ms	247 ms	287 ms	Free Search
	244 ms	250 ms	299 ms	Free Search (control)
	246 ms	261 ms	273 ms	Avoid Search
Saccade Planning Time (off-target)	241 ms	251 ms	292 ms	Avoid Search (control)
	199 ms	241 ms	290 ms	Free View
	208 ms	236 ms	310 ms	Free View (control)
	260 ms	275 ms	284 ms	Free Search
	260 ms	266 ms	310 ms	Free Search (control)
Fraction of trials where the ROI was visited within first 2 fixations	297 ms	297 ms	299 ms	Avoid Search
	290 ms	282 ms	310 ms	Avoid Search (control)
	89%	65%	8%	Free View
	87%	53%	7%	Free View (control)
	24%	32%	12%	Free Search
2-AFC Accuracy	33%	34%	6%	Free Search (control)
	28%	37%	10%	Avoid Search
	28%	29%	9%	Avoid Search (control)
	82%	78%	85%	Free Search
2-AFC Accuracy	79%	78%	80%	Free Search (control)
	72%	72%	82%	Avoid Search
	69%	72%	79%	Avoid Search (control)

Table 5 – Summary of timing and accuracy results for the 4 experiments

Experiment 1 - “Free View”, experiment 2 - “Free View (control)”, experiment 3 - “Free Search” and “Avoid Search”, and experiment 4 - “Free Search (control)” and “Avoid Search (control)” and the 3 entities (faces, text, cell phones).

Discussion

The results of the first experiment show that faces and text have a profound effect on the allocation of eye movements. The eyes are rapidly and strongly attracted to both, as compared to the slower and fewer fixations to cell phones (Figure 59). Faces and text rapidly attracted gaze - in over 65% of trials faces and text were foveated within the first two fixations. In contrast, only 27.3% of cell phones were foveated even after 7 fixations. Our results suggest similar mechanisms for attentional deployment to both faces and text, with a higher emphasis on faces. A second experiment controlling for the relative size and background of the images shows that faces are 1.49 times more likely to be attended to than text and 5.18 times more likely to be attended than cell phones.

In a third experiment, a separate group of subjects were asked to detect a small cross in the same natural scenes. The cross was present in half of the images. In 3 blocks, subjects were told before the trial began that the cross would not be present on the faces, text or cell phones (“avoid”), while in the other 3 blocks, the cross could be found anywhere (“free search”). We reasoned that as subjects were under time-pressure, they would not look at faces when they knew that the target was not located on faces, as this would be inefficient (same for text and cell phones). However, we see no such trend (Figure 60b and Table 5). There are no highly significant results discriminating between the avoid task and the free search task inferred from the fraction of fixations landing in the ROI. This was tested by comparing both the first one and the first two fixations of each trial. This observation may have not been seen due to the

instructions, forcing the subjects to think of the targets to be avoided, and thus causing them to look at those targets more often (Wegner, 1994).

Our data shows a consistent relationship between saccade-planning time and the percentage of trials in which the region of interest was attended during the first fixation. In the free viewing task we see that across entities saccade planning time is fastest for faces, second fastest for text and slowest for cell phones. This is consistent with the percentage of trials in which saccades landed on each of these entities' regions of interest, showing an inverse relationship to saccade-planning time. Further changes across categories also show this relationship. The fraction of trials in which faces and text are attended to initially is significantly lower in the free search task than in the free viewing task, this is reflected by the increase in saccade-planning time across the two tasks. For cell phones, there are no significant differences in saccade planning time across tasks, and we see no significant differences in the fraction of trials as well. We show that faces and text are significantly more salient than cell phones, by postulating that as the number of subjects who look at an ROI and the speed at which they look at it increases, the more salient it is regarded.

We report that it takes longer to initiate saccades for faces and text in the free search paradigm than in the free viewing paradigm, but we see no such difference for cell phones. This indicates that the sensitivity to faces and text in the search paradigm has been depressed, resulting in slower saccade times ($p < 10^{-6}$ free view vs. free search faces; $p < 0.01$ free view vs. free search

text), which is consistent with the decreased fraction of saccades landing in the regions of interest. This suggests that the salience of faces and text is task-dependent. However, this sensitivity towards faces and text does not vanish completely, as saccade-planning time did not increase all the way to the level of cell phones ($p < 10^{-5}$ for both vs. cell phones; t-test). This suggests that attentional deployment to faces and text is regulated only partially by top-down mechanisms.

The avoid experiment gives further support to the claim that attention towards faces and text is in part reflexive (Bindemann et al., 2005). In the avoid task subjects are better-off not looking at any of the entities, however, occasionally the subjects fail to avoid them. These are on-target saccades in the avoid task, and their saccade-planning time is not statistically different from on-target saccades in the free search paradigm. This indicates no decrease in sensitivity to faces and text, even when subjects are explicitly instructed to avoid the entity. In contrast, our data shows that on average it took 37 ms longer for subjects to look *away* from a face in the avoid task than in the free search task. Similarly, it took 22 ms longer to look *away* from a text in the avoid task than in the free search task. This increase in saccade-planning time suggests that extra computational effort may be required to actively avoid looking at faces or text because there is a natural tendency to attend to these stimuli.

If attentional deployment to faces or text were a top-down process, saccades to these stimuli would likely require longer to initiate, compared to saccades driven by a bottom-up process.

However, we observe the planning of these saccades to be equivalent in duration in both search tasks. Additionally, if saccades to faces and text were top-down, we would expect that avoiding these stimuli would not increase the time it takes to look away from these image elements. In fact, if the saliency of these elements could be quickly modulated by top-down mechanisms, we would expect to see a drop in the time it takes to avoid faces or text. Instead, we observe the opposite, with slower top-down driven computations preventing rapid deployment of bottom-up attention. More so, we see that when avoidance fails, saccade planning takes the normal amount of time. This suggests that top-down influences have not had enough time to influence attentional allocation and thus bottom-up forces are mainly present. We show that while indeed faces can be avoided, there are aspects such as latency measures that still reflect the bias.

Further temporal analysis uncovered other facts (Table 5). Timing analysis of saccade initiation for the free viewing task indicates that saccade initiation is fastest for faces; slower for text, and slowest for cell phones. These results go hand in hand with prior data reported by Fletcher-Watson (Fletcher-Watson, Findlay, Leekam, & Benson, 2008) showing that latencies for saccades to faces during the first fixation in a free viewing task were between 100ms - 249ms. Interestingly, off-target saccade times are indistinguishable from on-target saccade times in each of the image categories - seemingly implying that subjects look away from a face faster than they look away from text.

A possible explanation of the observed global depression in saccade latency could be that subjects may be adopting different general strategies across the block of images for each entity. To test this hypothesis we ran a control experiment where the three entities were randomly intermixed in one long block. The saccade-planning times for the different entities were consistent with those observed in experiment 3 - indicating no adaptation across the course of the experiment. We pulled out the stimulus set that followed the “free search” instruction, to be sure that the subjects had absolutely no knowledge of the stimulus category to come. Again, we found no significant differences between this set and our original experiments, thus showing this global depression phenomenon is not due to blocks or knowledge of the stimulus category. We also tested the possibility that this global depression is an image confound, as the images tested were disjoint sets across the stimulus categories. By embedding our entities into the same background images, we eliminated disparities in the image background that could be influencing the off-target saccades. However, even in this set of images we still see this global depression of saccade latency based on stimulus category.

A more likely explanation can be seen by considering a possible statistical selection bias of off-target saccades. For a non-face object to be the target of a saccade, it must have a saliency level larger or close to that of the competing face target. Given, then, that a non-face target was attended to, the expected salience of this target will be higher if there is a highly-salient face in the image than if there is not. Thus during the face trials, only the high-saliency non-face objects are able to out-compete the faces for attention and it is this subset used for the timing analysis. The fewer number of off-target saccades in the face trials select only these highly

salient competitors. The larger number of off-target saccades in the text and phone trials also select out these highly salient competitors and more competitors that are less salient – thus bringing down the average. Under the stipulation that saccade planning time decreases with higher stimulus saliency, these stimuli will then have a faster saccade-planning time than for less salient stimuli (such as cell phones, and to a lesser extent, text, in the present study).

This selection bias also explains the global discrepancy in the artificially embedded image set. Consider two competing stimuli that are in the embedded image set, say a highly salient banana, and a moderately salient chair. When a face is embedded onto the image that contains the highly-salient banana, the banana will win the competition and will draw a saccade with a fast saccade-planning time. When a face is embedded onto the image containing the moderately-salient chair, the face will win the saccade. Thus, the saccade-planning time of off-target saccades will be the speed of the saccade going to the banana. When a text object is embedded onto the same images, both the banana and the chair will first attract a saccade. Therefore, the saccade-planning time of off-target saccades for text will be slower than for the face.

For further insight, imagine that the saliency of competing objects is drawn from a Gaussian distribution. If the competition with highly salient faces wins over the bottom 80% of these stimuli, then only the top 20% contribute to off-target saccade-planning times. If competition with less salient text items cuts off the bottom 60% of these stimuli, then the top 40% will contribute to the off-target saccade-planning times. This leads to overall slower off-target

reaction times for images containing text (the average of the top 40%) compared to images containing faces (the average of the top 20%).

The early fixations to faces and text during free viewing and search (even when told to avoid looking at faces and text) suggests that subjects are biased to look at these objects independent of top-down mechanisms that drive eye movements and covert attention (Honey, Kirchner, & VanRullen, 2008). The tendency to look at faces is so pronounced that it is even possible to decode which image is associated with which scanpath by using the exact location of faces in the image (Cerf, Harel, Huth, W, & Koch, 2008). This ability to decode which image corresponds to a given scanpath is a measure of how attractive various image features are, as they reflect lower across-subject variability in scanpaths.

Although the exact way in which attention to faces is implemented in the brain is unclear (Johnson et al., 1991), it is well known that a number of cortical regions are specialized for faces, in particular the fusiform gyrus (Kanwisher et al., 1997; Tsao et al., 2006). In contrast to faces, we can be almost certain that attention to text is not an evolved process. Instead, it is likely that text sensitivity is developed through learning. It is possible, however, that the development of text itself was influenced by factors in the brain that control attention, and it is these factors that explain why text is highly salient (Changizi, Zhang, Ye, & Shimojo, 2006). Recent studies in fact argue in favor of a specialized area in the brain for words and text – the “visual word form area” (Cohen, Jobert, Le Bihan, & Dehaene, 2004), which could take part in

the allocation of bottom-up attention demonstrated in our tasks. Interestingly, long term experience may also play a role in development of face recognition abilities in the brain (Golarai et al., 2007). Regardless, our results show very similar patterns in the way by which attention is allocated towards faces and text, suggesting similar mechanisms for attention deployment in the brain.

Further studies of the interaction between top-down and bottom-up mechanisms leading to the attention allocation to faces show that the ability to rapidly saccade to faces in natural scenes depends, at least in part, on low-level information contained in the Fourier 2-D amplitude spectrum (Honey et al., 2008). This suggests that a bottom-up saliency model incorporating image features may be able to account for part of the attention allocation.

Our experiment demonstrates that faces and text are very attractive and are difficult to ignore, even if there is a real cost associated with looking at them. It remains to be seen how the single neuron substrate of faces and text reconciles the sometimes conflicting demands of bottom-up and top-down inputs. Our improved model of saliency-driven attentional deployment is of relevance to a host of military, commercial, and consumer applications. The success of incorporating additional biologically-inspired detectors for high-level cues suggests similar attention allocation patterns to those used by the brain.



There's plenty of time at the bottom:

What happens in your brain
before a saccade is initiated

*There's plenty of room at the
bottom.*

Richard Feynman

Having shown that observers attend to important high-level objects in general, and faces and text in particular, rapidly, most of the time, regardless of the task, and in a bottom-up manner we wanted to investigate the nature of the urgency to do so. As we shown earlier that a visit to a face region is happening very quickly, sometimes before our conscious mind had enough time to thoroughly process the entirety of the image, we wanted to investigate the brain mechanisms which lead to such a rapid saccade. More so, we were interested in seeing whether we can infer from the latencies of the saccades something about the nature of the processing which led to the eye-movement.

The purpose of our follow-up research was to determine whether some information about the image could be processed without any conscious information? More so, given the long-lasting debate between the involvement of bottom-up and top-down information in guiding our attention, we suggest here a model of interplay between early subcortical and cortical mechanisms that drive our attention rapidly, with later involvement of high level frontal cortex regions that alter our attention allocation. We use the saccades that happen very early when an image is presented, prior to the involvement of conscious decision making in guiding our attention to study the mechanisms that make us attend to things before we know what they are, where they are, or why we want to look at them.

Introduction

In deciding where to look it is advantageous to be able to do so as rapidly as possible. However, in order to decide where to look we need information on what is present in our visual field. The delay involved in waiting for visual information to reach frontal cortex and effect a goal-directed saccade is too long to be useful in many situations. The brain needs simpler, faster mechanisms of directing eye movements. These are usually referred to as covert mechanisms, driven by bottom-up visual saliency (Itti et al., 1998), as opposed to the higher-order goal-directed top-down mechanisms. Bottom-up saliency is thought to be dependent on basic sensory inputs such as colour, orientation, intensity, and contrast. Following top-down mechanisms then direct our eye movements on the basis of cognitive strategies - towards food if we are hungry, or towards a single person out of the many in the picture, if he happens to be our son (A Oliva et al., 2003).

A second dichotomy exists between subcortical visual structures and cortical structures governing our vision (Shipp, 2004). Regions such as the superior colliculus (SC), along with other subcortical structures, are capable of initiating saccades to visual targets autonomously. However, while such structures may be able to do so, they lack any information as to *what* a target is, how interesting it is, or how much we want to look at it. In order to derive the form of an object it is necessary for incoming visual information to be processed by higher visual centres in our cortex. It is this necessity which is commonly regarded as the reason for saccadic latencies being as long as they are. While we can make a saccade sooner, we need to wait until we can infer *what* there is to look at before making an intelligent decision on *where* to direct our attention to.

We are, however, very rarely confronted with a totally new visual image. Our surroundings are commonly static and consist of little movement with no suddenly appearing targets. Despite this we continually make *spontaneous* or *scanning* saccades. Under these conditions our visual environment is utterly predictable, the timing and location of any saccade is known to us and each new visual image that arrives on our retina is effectively known ahead of time. Where this is the case it makes no sense to wait for costly cortical analysis of each new visual image, if we know ahead of time where an object of interest may be located then it would be prudent to have a separate mechanism capable of directing an eye movement to such an object without waiting for the cortex to decide.

There are two ways of doing this: either speed up cortical decision mechanisms, or let subcortical mechanisms drive the next saccade. A Linear Approach to Threshold with Ergodic Rate (LATER) model provides a means by which both can be done (Carpenter & Williams, 1995; BAJ Reddi, Asrress, & Carpenter, 2003; BA Reddi & Carpenter, 2000). The LATER model proposes that a detection signal S rises linearly at a rate r until some threshold level is reached, at which a saccade is triggered, with latency T . If r varies randomly from saccade to saccade with a Gaussian distribution then the result will be a latency histogram with a tail skewed to longer latencies. More specifically, the saccade latency can be reflected as a straight line on distribution plotted cumulatively on a probit ordinate and reciprocal abscissa.

Additionally, the LATER model represents an ideal Bayesian decision making process. The decision signal S represents the log likelihood of the hypothesis being correct at any given time. With an initial value S_0 representing log prior probability and a linear rise until confirmatory

information is reaching a threshold S_r . This threshold reflects the probability which justifies the initiation of a saccade.

$$\frac{1}{T} = \frac{r}{(S_T - S_0)}$$

Experimental manipulation of prior probability, through alteration of stimulus predictability, alters the amount by which the decision signal needs to rise until the threshold is reached. Such manipulation leads to a swivelling of the main distribution around its origin, as predicted by the LATER model (Carpenter & Williams, 1995). Thus an increase in prior probability, in response to a predictable stimulus, results in faster saccades.

Additionally, plotting the distribution of saccadic latencies on a reciprobbit plot reveals a subpopulation of *early saccades* (ES). These saccades make up a small fraction of the total and usually lie on a separate straight line that passes through 50% at infinite time (Figure 61). This can be modelled as a maverick LATER unit, with a mean rate of rise of 0, such that 50% of the time it initiates no response, in parallel with a main LATER unit, and with a very large variation in rate such that occasionally it overtakes the main unit and fires a saccade first. This early population is thought to represent relatively automatic responses, created by a subcortical structure such as the SC.

We tested the existence of such early saccades using eye-tracking data of subject's viewing images of natural scenes, providing a target rich environment, in a task that is indifferent to the speed of viewing. Subjects were instructed to look at images and judge their interest. We tested the proportion and latencies of the ES and propose an explanation for differences in latencies based on various pathways in the visual system.

Owing to prior studies showing that faces and text act as strong attractors for fixations, we were able to establish a controlled frame even in a packed natural scene environment. This enabled us to specifically test the relative saliency of faces and text in these images. The reciprobbit plot provides an excellent means of separating out two populations of saccades - main and maverick - according to saccadic latency on a subject-by-subject basis. We could therefore individually analyse the saliency of faces or text to these two apparently independent mechanisms of saccadic generation.

Our results suggest that subcortical structures are capable of generating rapid, accurate saccades to salient targets in predictable images. They show a separate, cortical source of bottom-up saliency to objects within a visual scene which disappears within a few fixations. Finally they provide strong evidence of “race-to-threshold” competition between visual objects in a manner described by the LATER model, while demonstrating rapid alteration of target prior probability over single fixations. To this effect we additionally propose to amend the definition of bottom-up attention to account for two separate processes by which bottom-up attention is directed: subcortical bottom-up and cortical bottom-up.

Methods

Nineteen subjects viewed a set of natural scenes images (1024 x 768) in two experiment types. In the first experiment (“free viewing”) images were presented to the subjects for 2 seconds, after which they were instructed to answer the question “how interesting was the previous image?” using a scale of 1-9 (9 being the most interesting). Subjects were not instructed to look at anything in particular; their only task was to rate the entire image. In the second experiment (“search task”) subjects first viewed a probe (face or an object) for 600ms and then saw an image which either did or did not contain that probe for 2 seconds, after which they were instructed to answer the question “did the probe appear in the previous image? (y/n)”. Fifty percent of the images contained the probes, and 50% of the probes were faces.

Out of the entire set, 344 images contained one or more faces, and 80 images contained one or more discrete text element. The face images were photographed in indoor and outdoor environments. The images included people of various skin colours, ages, and postural positions. A few images had face-like objects (smiley T-shirt, animal faces, masks, faces carved in stone, etc.). The text images were taken from the Internet and were chosen such that only a few text objects appeared in the scene. Some of the images contained objects such as a colourful Rubik’ cube, a toy fire truck, plastic banana, etc. These flashy salient objects served as competitors for the attention of subjects viewing the images with the faces or text. The average sizes of the faces/text/objects were $4\% \pm 1.8\%$. All images were chosen as images that are representative of a real-world scene. All the images and the subjects’ scanpaths are taken from www.fifadb.com (Cerf, Frady, & Koch, 2009).

Image order within each block was randomized throughout the experiment (see also Cerf et al., 2009 for details on the experimental design). Subjects viewed the images in 3 blocks of 200 images. Some images were repeated within the image block, but altogether, each subject saw 151 unique images in the entire experiment.

Prior to each session, the subjects' gaze was determined through a calibration process. Before each stimulus onset, the subjects were instructed to look at a black cross at the centre of the screen. If the calculated gaze position was not at the centre of the screen, the calibration process was repeated to ensure that position was consistent throughout the experiment. Eye-position data was acquired at 1000 Hz using an EyeLink 1000 eye-tracking device (SR Research, Osgoode, Canada). The images were presented on a CRT2 screen (120 Hz), using Matlab's Psychophysics and eyelink toolbox extension (Brainard, 1996; Cornelissen et al., 2002). Stimulus luminance was linear in pixel values. The distance between the screen and the subject was 80 cm, giving a total visual angle for each image of 28° x 21°. Subjects used a chin rest to stabilize their head. Eye movement data were acquired from the right eye alone. All subjects had normal or corrected-to-normal eyesight. All subjects were naive to the purpose of the experiment. The experiment was undertaken with the understanding and written consent of each subject. All experimental procedures were approved by Caltech's Institutional Review Board.

Fixations were determined by the built-in software of the eye-tracking system. The "initial fixation" is the fixation occurring before stimulus onset, when the subjects are focusing on a centred fixation cross, and is not counted as part of the ordered sequence of fixations.

The fixation latencies for each subject were plotted in a cumulative histogram on a probit scale. Given that the latencies distributions for all images were typically skewed Gaussians, this resulted in a straight line with a few early maverick fixations that landed outside of the linear fit. We fit these maverick fixations separately and separate the two types of fixations at 50% probability (see Figure 61 for an example of the reciprobbit plot for two subjects). The fixations within each fit are termed early fixations and late fixations.

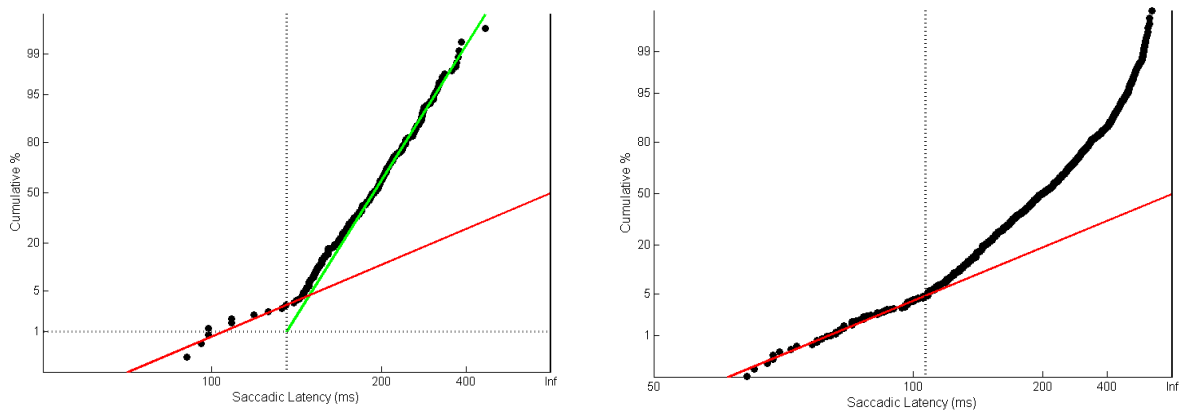


Figure 61 - Examples of saccadic latency fits of two subjects

The latencies of the saccade onset of two subjects (chosen arbitrarily) viewing 344 images were fit to a linear Gaussian on a cumulative plot. The saccades latencies are clearly separable to two distinct distributions. We separate these fixations and term them early ('maverick') and main distribution saccades.

Left: The early distribution is split on the basis of deviation from the main distribution.

Right: the main distribution is split on the basis of deviation from the linear fit of the early distribution.

We calculated the fraction of fixations in a “region of interest” (ROI). We manually defined the ROIs as a minimally sized polygon around each face, text or object in the images, creating a binary heat-map describing the location of the entities. We calculate the fraction of fixations in an ROI by summing the number of fixations that fall within this region over all trials and then dividing by the number of trials. A trial consists of one subject being presented one image and will contain one ordered sequence of fixations.

To compute chance level of performance in our analysis, we calculated the ratio of the fraction of fixations that land in a target’s ROI to the fraction of a baseline distribution. The baseline for a particular image is the fraction of all subjects’ fixations from all other images that fall in the ROI of the particular image (for detailed explanations see Cerf et al., 2009). This takes into account the varying size and locations of the ROI in all images (as these factors both influence how likely a certain region is to be fixated on by chance) and the double spatial bias of photographer and observer. As the baseline value estimates the chance level of landing in a given ROI, we make a comparison with this baseline value to the data by dividing the fraction of fixations landing in the ROI by its baseline value. This allowed us to indeed test the fraction of fixations from various latencies being drawn to a specific location in the images.

Results

Free viewing

Early Saccades

The clear distinction made between the main distribution of saccades and the early distribution makes the two populations amenable to individual investigation. Initially we looked at the first fixation subjects made after image presentation; meaning they were presented an unpredictable visual field containing multiple targets in which there were either no, one or more faces. As subjects produced reciprobts with varying mean and slope of latency distributions we first normalised the latency, where a latency of 0ms represents the cut-off between the early and main distributions, before placing the fixations into 20ms bins on the basis of their normalised latency. We then calculated the percentage of fixations in each bin which landed on a face and thus created a histogram of normalised saccadic latency against percentage of fixations landing on the face: a saliency histogram (Figure 62).

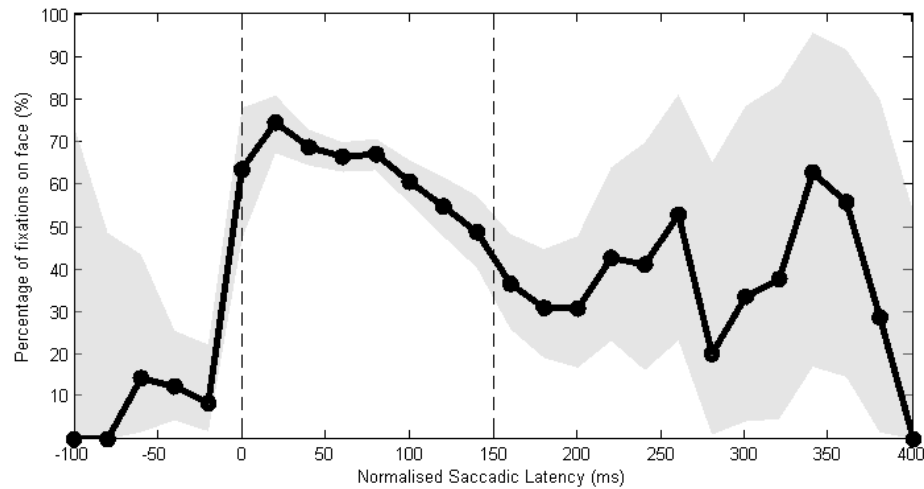


Figure 62 - Percentage of early saccades landing on a face region of interest

Dashed lines mark the boundary of early (<0ms), middle (0 - 150ms) and late saccades (>150ms). Latencies are normalized to the breakpoint between the two saccadic distributions (see Figure 61). While the middle saccades land on the face in between 48.5 and 74.2%, our subjects visit the face in their first fixations only 8.47-14.3% in their very early saccades. The shaded area marks the 99% confidence intervals.

We found a significant increase in the percentage of fixations landing on the face at a normalised latency of -10ms (that is, 10ms prior to the time set as the onset of the initiation of later saccades). This increased proportion of facial fixations is maintained throughout the main distribution, though it shows a decline after 100ms. ES show an attraction to faces that is not significantly above chance and significantly lower than main distribution saccades ($p < 1.3 \times 10^{-45}$, Wilcoxon rank-sum). The increased proportion of facial fixations at -10ms is attributed to the ES of the main distribution being below the cut-off latency. On this basis we amend the cut-off point to -10ms to maintain a 'pure' distribution of early saccades. There are two facets to the

main distribution: an early peak in facial saliency, followed by a general decline in facial saliency with increasing saccadic latency. Thus, we define a distinction between the two at +150ms normalised latency. We call the earlier population “intermediate” saccades (IS), while the longer latency population are “late” saccades (LS).

We looked at the properties of these distributions as a group showing their relative attraction to faces and text. Comparing fixations to faces and text viewed by subjects we see that fixations to text-containing images, which are fewer in number with fewer subjects, show a pattern similar to that of faces. There are no ES made to text, while on the main distribution there are frequent fixations on text. Intermediate saccades show high text and facial saliency while LS goes less to both (Figure 63). As it is extremely unlikely that we have a subcortical text detector we can assume that all saccades on the main distribution are cortical in nature and, more specifically, have access to high level “what” information (Van Essen, Anderson, & Felleman, 1992). These results suggest that these early saccades are generated by a separate mechanism which does not have access to such information.

While the first fixation is made to an unpredictable image, the second fixation is made to an image which is entirely predictable, the information about the environment has already been processed by our brain, and we are aware at least of the gist of the content in the visual field. During a saccade there is a period of rapid retinal slip before a new retinal image appears on its completion. As we generate the saccade the timing and location is known, and thus so should the properties of the new retinal image including the location and identity of visual objects. To look at how this affects saliency we performed a reciprobital analysis of the second fixation (Figure 63b). Given the onset of a new visual image at the end of the saccade onto the first

fixation, the earliest this new retinal image can be acted on by cortical mechanisms is the beginning of the main distribution. Prior to this however it is possible that saccades could be generated by cortex in a corrective manner in response to the first visual image; an *open-loop* cortical saccade. To test which of the two pathways led to the initiation of the second saccades we looked at a fourth population of saccades, those generated within 60ms of fixation onset, and thus generated *before* it is possible for the new retinal image to be acted on at all. This assumes that the distance between the retina and our visual systems poses a minimal timing on information accessing, giving a lower boundary of roughly 60ms for initial information access (Busettini, Masson, & Miles, 1997). We call this population “very early saccades” (VES), and it can only contain saccades generated in response to the previous retinal image.

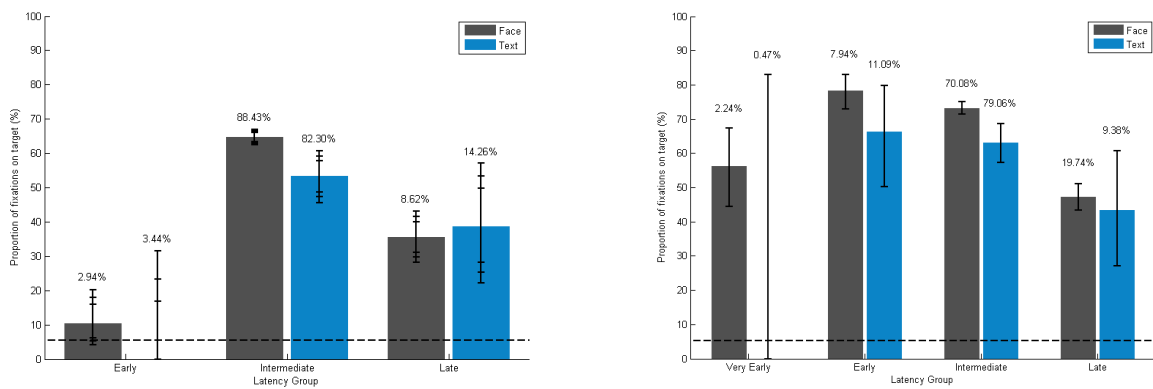


Figure 63 – The proportion of fixations landing on faces/text

Calculating the average proportion of fixations on the face (see figure 2) or text targets (left) shows a significant difference between the 3 groups (ES, IS and LS) for either target types. Gray bars marks the proportion of first fixations to land on face ROIs, while blue bars are text ROIs. The error bars are: left; 95/99/99.9% confidence intervals, Right; 99% CIs. Horizontal

dashed black line is chance. Above each bar are the proportions of fixations which make up this saccades category. Breaking the saccades following a prior fixation (second fixations/saccades) to four groups (right) we see a large increase in the proportion of fixations landing on face, over those landing on text. The increase is seen mainly in the very early saccades, which can be explained as correction saccades for an over/under-shoot of the fixations based on the new retinal information received from the new image formed after the initial fixation. The difference between all group distributions is significant, with the biggest difference for the early saccades.

VES represent exclusively open-loop saccades and show a considerably higher than chance attraction to faces with 55% of saccades falling on faces. However ES, with a latency longer than 60ms, which are capable of being produced in response to the new retinal image, show a significant increase in facial saliency compared to VES ($p = 2.4 \times 10^{-7}$, Wilcoxon rank-sum). There is also a significant increase in the number of saccades, with four times more early saccades than very early saccades ($p < 1 \times 10^{-20}$, Wilcoxon rank-sum). This increase in saliency and number occurs in response to new visual information, but happens before the cortex is capable of discerning semantic information from the new retinal image.

Inter-saccadic changes in LATER units

In order to analyse the inter-fixation changes better we took data from a single subject and then, using only saccades which landed on the face, broke it down into individual fixations. Our data sets therefore correspond to LATERian face units on individual fixations. We used these to plot reciprobits for each fixation separately (Figure 64).

As this corresponds to a single subject we can consider the variance of the unit to remain constant. As such the *intercept* with the $T = \text{infinity}$ axes becomes a measure of the *mean rate of rise*. The higher the *intercept* is, the higher the *mean rate of rise* of the face unit. The gradient becomes a measure of prior probability; the less steep the gradient the higher the expectation for a face; the higher the prior probability. Looking at the prior probability of the face unit we see that for the first fixation the image is unpredictable and the prior probability of the face unit is relatively low. However, for the second fixation onwards the prior probability becomes higher - the gradient of the line decreases - as the image becomes entirely predictable. This difference is significant across all subjects.

This change in prior probability is coupled to the emergence of early saccades which land on the face. Crucially, these early saccades still lie on an obviously separate trend line which passes through 50% at $T = \text{infinity}$; they are still maverick saccades. The trend lines shown are constrained to pass through this point and demonstrate swivel around this point, corresponding to changing prior probability (BAJ Reddi et al., 2003).

From fixation two to three we see a very distinct rightward shift in the reciprobity which corresponds to a change in the mean rate of rise independent of any change in prior probability or variance. There is a larger change in the intercept - the mean rate of rise - from fixation one to two.

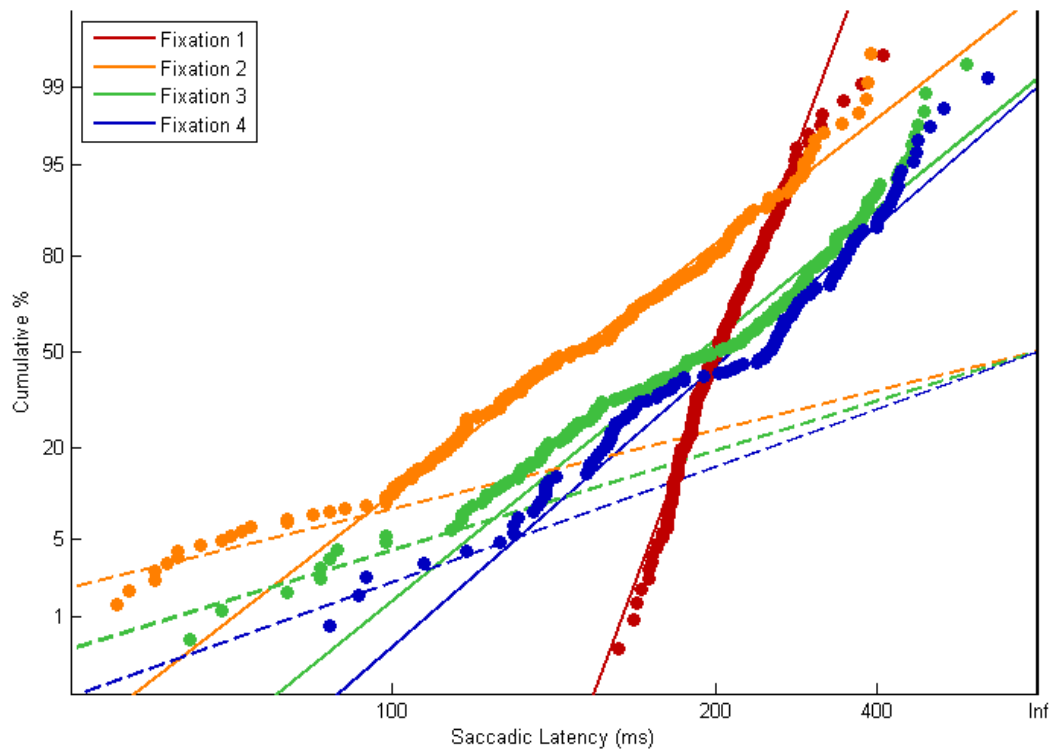


Figure 64 - Fixations to the faces reflect increased information and mean rate of rise

A comparison of the first to the fourth fixation of a single subject comprising only saccades which landed on a face. Solid lines represent trendlines of the main distributions, dashed lines are trendlines of the early distributions. We see a significant ($p < 10^{-10}$, Kolmogorov-Smirnov) difference which is attributed to a swivel in the distribution for the first fixation relative to the second, third and fourth. There is additionally a dramatic change in the y-intercept from the first fixation to the second fixation before settling by the third and fourth fixations

LATERian Race-to-Threshold competition

There is no change in the amount of sensory information being provided between fixations so the changes in the mean rate of rise of the face unit must be representative of some other process. Equally, LATER units representing different visual objects may also have different mean rates of rise. With this in mind Figure 62 shows behaviour that is consistent with a LATERian race-to-threshold mechanism of saccadic decision. We can explain the behaviour of facial saliency in the main distribution as a feature of LATERian competition between multiple visual objects of differing mean rates of rise.

Consider simply a two unit system of a face unit and a “background” unit. If the mean rate of rise of the face unit is higher than that of the background unit then an earlier saccade is more likely to be generated by a face unit. Conversely the slower the rate of rise of the face unit is the more likely it will be overtaken by the “background” unit and thus as latency increases the proportion of fixations on the face decreases. So we can think of the change in proportion of fixations on the face with saccadic latency to be related to the difference in the mean rate of rise of a face unit against the rest of its environment.

With LATERian competition in mind, changes in the mean rate of rise of objects should manifest as changes in the saliency histogram. Given the changes in the mean rate of rise from fixation to fixation, we show that the saliency histogram for the first four fixations on face-containing visual images with 99% confidence intervals changes with time, becoming progressively flatter. The difference between the saliency histograms for each fixation is significant (Figure 65a).

In order to better understand how these change in the mean rate of rise impact the shape of the curves in figure 5a we modelled LATERian competition between a face unit and an object unit (Figure 65b). We confirm that when the face unit has a higher mean rate of rise than the object unit the proportion of fixations on the face starts high and decreases. We find that the larger the difference in the mean rate of rise between the face unit and the object unit the steeper the change in proportion of fixations landing on the face is. Conversely the closer the mean rates of rise of the two units are the flatter the saliency histogram.

Rather than looking at individual results we can take population data by calculating the y intercept and gradient of the reciprobits for each subject and then taking the population mean gradient and y intercept to create an average reciprobit for the population. At the same time we can do this not just for the face unit, we can do it for the object units using the previous objects. Figure 65c shows the results for the whole population, with the data for both faces and objects. The first fixation shows low expectation and high mean rate of rise for both faces and objects. The expectation then alters significantly and remains fairly constant for all following fixations, while the mean rate of rise drops progressively until the third or fourth fixation in both faces and objects. This is in agreement with the individual data.

Interpreted in conjunction with Figure 65b, Figure 65c explains the saliency histograms in Figure 65a. There is a large difference in the mean rate of rise between faces and objects for the first fixation, thus creating a steep drop in the proportion of fixations on the face with increasing latency. For the second fixation, while the mean rate of rise of both the face and object units has dropped, the difference between the two has decreased, thus the change in the

saliency histogram decreases. By the third and fourth fixation there is almost no difference in saliency between faces and objects and thus there is a correspondingly flat saliency histogram.

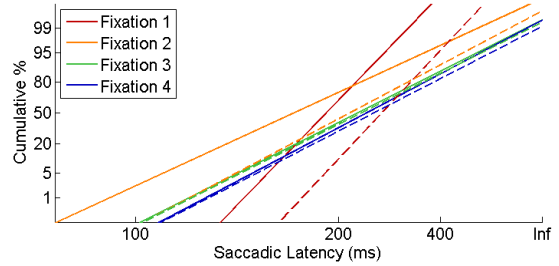
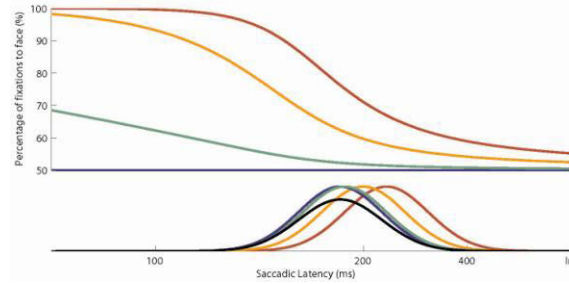
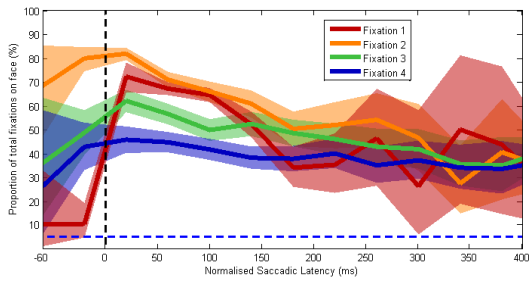


Figure 65 - Competition and mean rate of rise affect the saliency histogram

LATERian competition and differences in mean rate of rise determine the shape of the saliency histogram. Calculating the proportion of fixations on the face as a function of normalised saccadic latency for fixations one through four. Latencies are aligned to the cut-off point between early and intermediate saccades as in Figure 62. While the first fixation (red solid line, and red shaded area for 99% confidence intervals) shows a very low percentage of face fixations in the early saccades and a significant increase towards the intermediate saccades and a gradual drop towards the late saccade we see a different distribution for the later fixations (2-4). Later fixations initially have a very high proportion of fixations to the face (as shown in

Figure 63b), with a gradual drop in the proportion towards the intermediate and late fixations.

All 4 distributions show similar values for the later latencies.

Top right: modelling the effect of differences in the mean rate of rise between two competing LATER units (bottom part of the panel; black distribution represents the face unit) on the estimated proportions of fixations on the face (upper part of the panel). The smaller the difference in the mean rate of rise between distributions, the flatter the saliency histogram becomes.

Bottom: Comparison of fixations to the faces (solid lines) and to the background (dashed lines) for the first to fourth fixations. We see a difference in the mean rate of rise between the first and the following fixations altogether, and a significant difference between fixations to the face and the background. The differences in the y-intercept between the face and background reflect a weighting of incoming information – which we attribute to the saliency of objects – generating the differences in the mean rate of rise and thus the latencies as shown in the above panel. Faces are seen as more salient altogether and are viewed at a higher proportion in faster saccades.

Top-Down Influences

In order to test the top-down influences on saccadic decision we used a search task, where subjects were given a clear information and directive as to where to address their viewing in the scene. Subjects were directed to look at a face or other object prior to image onset. Since faces are a salient attractive object, they ended up viewing the faces often even despite the directive. We quantified the proportion of fixation on the face, separated by the two directives given (find

an object vs. find a face), and the availability of information on the image (first fixation - no information, and later fixations with availability of prior information).

Faces are visited in the search task less than in a free viewing; however they are still visited much more than chance (Figure 66a).

However due to the nature of the task we cannot compare the results here to those in a free-viewing task as reciprobts. Instead we looked at goal completion (Figure 66b) where we looked at fixations to the face during a face search task separated according to whether the face had previously been fixated or not, i.e. the goal had been completed or not respectively. What we see is a rightward shift in the curves with goal completion. More specifically the mean rate of rise of the face unit remains high until the face is fixated, at which point the goal of the task is completed and we see a corresponding decrease in the mean rate of rise of the face unit more in line with previous results (Figure 64).

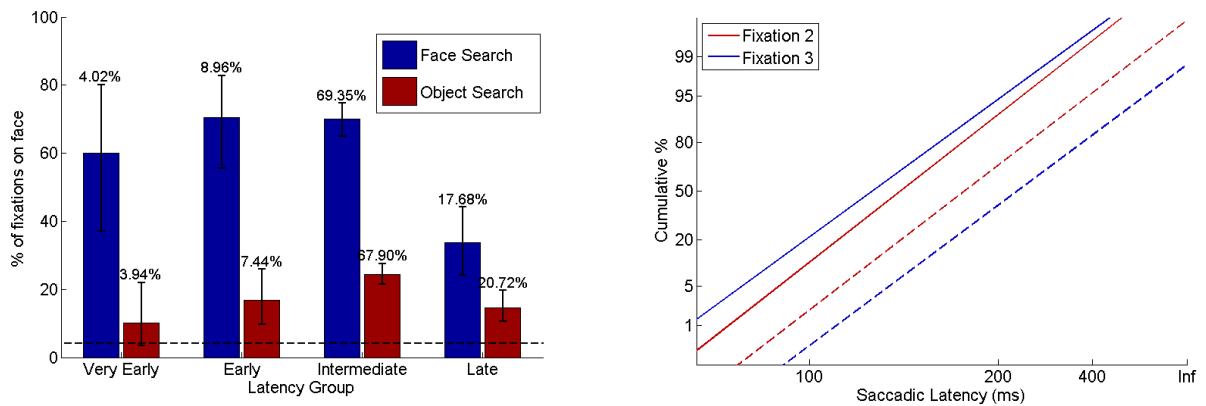


Figure 66 - Proportions of fixations on the face in the search task

We measured the proportion of fixations to the face regions of interest when subjects were instructed to find the face in an image (blue bars) and when the subjects were instructed to find an object (red bars) during the second fixation. Subjects show a significant decrease in the viewing of faces when they are instructed to locate an object in a short time, suggesting that top-down influence can affect even the early saccades. When subjects are instructed to find the face they are increasing their prior probabilities for face finding, thus increasing the proportions of face visits in all saccade latencies categories, including the very early ones.

Left: Top-down attribution to the face fixations. Generating the linear fit for the main distribution of each of 19 subjects facial fixations during a face search task we averaged all the slopes and intercepts and generated an average linear fit for all subjects' latencies populations. Taking only the second and third fixations, we show the latencies of fixations that went to the face after it was visited in an earlier fixation (dashed lines; goal previously completed), and the ones where the face was not visited earlier (solid lines; goal not previously completed). When the subject had viewed the face earlier we see a decrease in the rate of rise, which - according to the LATER model - is attributed to change in the amount of information, but here is a quantifier of top-down effects.

Discussion

Our results show that based on the early latencies of saccades in free viewing of natural scenes, the eye-movement to target in rapid viewing are the outcome of an interplay between cortical and subcortical mechanisms. While the eyes are strongly and rapidly attracted to faces and text, we tend to visit those only once cortical information is updated. Thus, the viewing of a natural scene should be broken to that governed by bottom-up saliency, its effect of early saccades, and the interaction between brain structures to control the saccade in a scene. While the free viewing task enabled us to tap into the underlying competition between cortical and subcortical regions in the control of our saccades to high-level objects, we used the search task to learn about the later competition between top-down and bottom up driven attention to control our gaze (Connor, Egeth, & Yantis, 2004; Ogawa & Komatsu, 2004).

Bottom-up Saliency

Prior studies suggests that the mean rate of rise of a LATER unit is a function of the rate of arrival on incoming confirmatory sensory information (BAJ Reddi et al., 2003). However in these results we see changes in the mean rate of rise between objects and between fixations. From fixation to fixation there is no change in the arrival of incoming sensory information. This led us to a reinterpretation of the LATER model. Rather than taking the mean rate of rise of the decision signal to correspond to rate of arrival of confirmatory sensory information we can consider it to correspond to the saliency of the visual object. This includes the original interpretation, whereby an object's saliency weights incoming information, and thus would be confounded in previous experiments. We can interpret the observed large initial mean rate of rise of object units, which rapidly decays over a second, to correspond to bottom-up saliency.

This is supported when we look at the attractiveness of four objects (toy banana, toy car, Rubik cube and phones) against the rest of an image in images either containing a face or no face during the second fixation (Figure 67). As we have already seen, the relative difference in the mean rate of rise, and thus the saliency, of visual objects determines the shape of the saliency histogram. In images containing faces we see the reciprocal pattern of Figure 62 as saccadic latency increases, the proportion of fixations landing on the object increases. This fits with an interpretation that the face is relatively more salient than the objects and thus a higher proportion of lower latency fixations will be generated by the face unit. If the face unit, by chance, starts with a slower rate it is more likely to be overtaken, thus an increasing proportion of longer latency saccades go to the objects.

However when there is no face present in the image the object is now relatively more salient than the rest of its image and shows the same pattern as face fixations; decreasing in proportion with increasing saccadic latency. One further point that can be taken from Figure 67 is that early saccades are also capable of targeting objects. However, they are far more selective for the most salient object in a visual scene than main distribution saccades; when a face is present fewer early saccades are generated to an object than main distribution saccades.

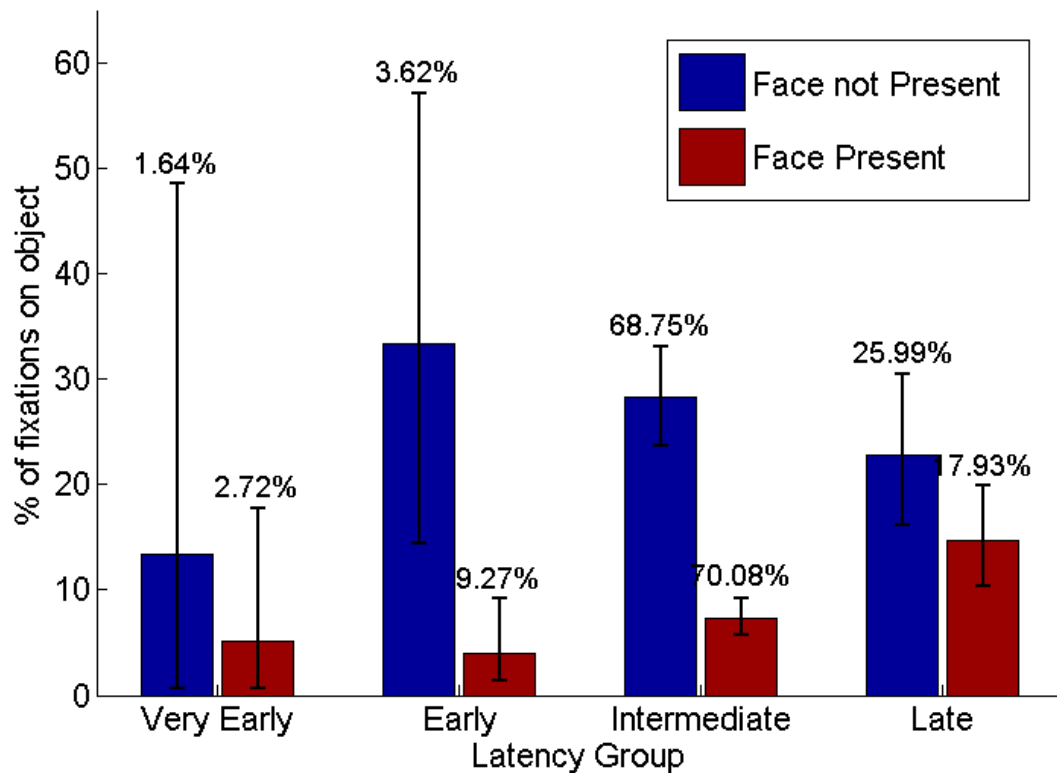


Figure 67 - Differences in relative saliency affect fixation location

We measured the proportion of second fixations which landed on one of four objects (toy banana, toy car, Rubik cube and phones) during a free-viewing task of images either additionally containing a face (red bars) or not containing a face (blue bars). When an image contained a face there were fewer fixations on objects at all latencies than when an image did not contain a face. When a face was present, and thus was the most salient visual object, the proportion of fixations landing on objects increased with increasing latency. When there was no face present, and thus the objects were relatively more salient than their background, we see the proportion of fixations to objects decreasing with increasing latency.

Top-Down Influences

This however is an incomplete picture as we have also seen that the presence and completion of a goal can also have effects on the mean rate of rise of LATER units. In Figure 66b we see the normal trend for bottom-up saliency, a declining mean rate of rise of the face unit, taken over and halted by top-down mechanisms until the goal of fixating on the face is achieved.

As Figure 65 and Figure 67 demonstrate, in LATERian competition the relative values of the mean rates of rise of competing units has a dramatic effect on the proportion and latency of fixations an object expects to receive. The unit with the highest mean rate of rise gains the largest percentage of fixations, particularly among the fastest saccadic latencies. Bottom-up saliency then provides a means by which the sensory information encoding the most relevant objects are weighted such as to increase the mean rate of rise of the relevant LATER unit and generate a saccade to those objects quickest and most often.

The presence of a goal then provides top-down influences for the same purpose, by raising the mean rate of rise of a particular LATER unit we increase the proportion and decrease the latency of saccades targeted to that goal. Given these results and the nature of the mechanism we hypothesise that the mean rate of rise is in effect a measure of a *utility* signal. A complete decision mechanism requires a consideration of expected utility whereby a decision signal is scaled to reflect associated utilities.

While our tasks may be simplistic, both bottom-up saliency and the changes associated with goal-completion represent choices based on utility. Fixating faces first in a scene is useful thus faces have a high initial saliency which declines rapidly. In a face search task fixating a face is

our only goal and thus its mean rate of rise is kept high until it is fixated, the goal completed, and then allowed to fall again.

Early Saccades

In contrast to this bottom-up saliency, early saccades represent a separate mechanism for driving saccades to a visually salient target even more rapidly. ES are distinguished as functionally distinct to cortical saccades on the basis of the reciprobbit plot, where they are seen to have a different distribution to the main distribution. Carpenter suggests these saccades can be modelled as a maverick LATER unit with a mean rate of rise of 0 and a large variation. While the main distribution varies in terms of shift and swivel, the distribution of ES does not shift but only swivels about its intercept. This suggests that the mean rate of rise of these maverick LATER units doesn't change, but the expectation of objects does.

ES on the first fixation are rare, and do not respond to semantic content. However on subsequent fixations early saccades are generated with considerably higher frequency and they target objects on the basis of semantic information; landing on faces and text. This is not simply a matter of prediction in the cortex speeding up its own analysis; the ES are still clearly maverick saccades and functionally distinct to cortical main distribution saccades. This suggests that while early saccades may themselves not be capable of processing information about form, the mechanism can be "taught" what visual targets are important in a predictive manner, thus allowing rapid, accurate scanning saccades to salient targets without having to reprocess the entire new retinal image to a high level. Indeed our results show swivel around the intercept of the early distributions suggesting alteration of the prior probability of these LATER units.

As these results also suggest that ES are not simply open-loop cortical saccades, given their increased proportion and accuracy compared to VES, that ES are generated by a functionally distinct mechanism which does not have access to “what” information as is available to cortical mechanisms, and given their extremely short latency we suggest that these results support the interpretation that early saccades are generated by subcortical mechanisms, such as the superior colliculus (SC).

Latencies in the visual system

Figure 68 and Table 6 summarise the known latencies of the visual system. ES are generated as early as 60-100ms after the onset of the new retinal image. The short latency already explains why this mechanism lacks access to “what” information about the retinal image as it takes visual information at least 90ms just to reach the higher areas of the ventral stream. This leaves two possibilities for the location of the mechanism behind ES; subcortical generation in the SC or generation by cortical mechanisms without higher level information processing.

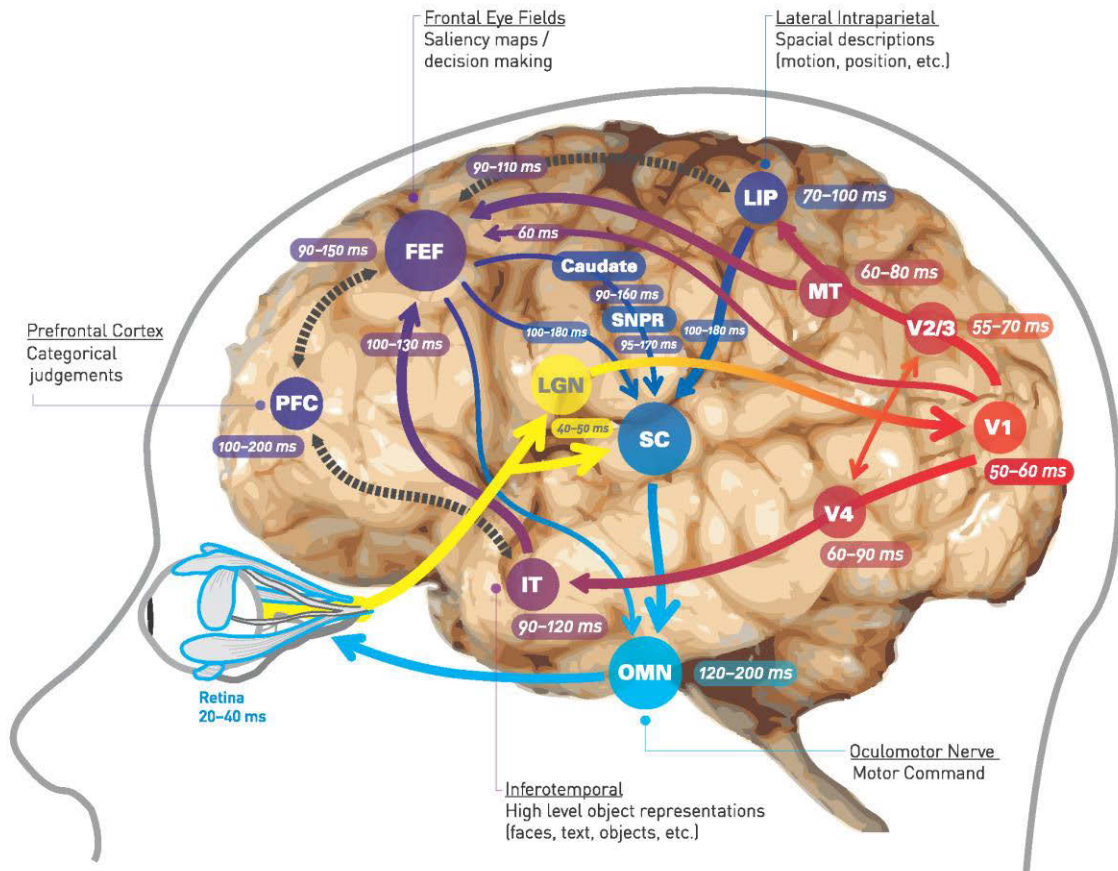


Figure 68 - Latencies in the visual pathway

Humans saccade to the target quickly, with retinal movements that average 120 to 250ms after image onset. Depicted is a plausible route between the retina and back via the oculomotor nerve which guides the eye towards its target. Information from the retina is relayed by the lateral geniculate nucleus of the thalamus (LGN) before reaching V1, the primary visual cortex. From there, processing continues through the dorsal pathway, the *where* pathway, through areas V2, V3 to the lateral intraparietal region (LIP) where special information is processed, and through the ventral pathway, the *what* pathway, through V4 to the inferior temporal cortex

(IT), which contain neurons that respond specifically to certain objects. Both areas IT and LIP project, with different latencies, to the frontal eye-field (FEF) which waits for information from both before forwarding it. Due to the very early saccades we show in our data, we posit an additional direct projection between V1 and FEF which generates the very early saccades. Area IT projects to a variety of areas, including the prefrontal cortex (PFC) which provide high-level cognitive judgments to the future saccade planning. Information from FEF is then forwarded to the superior colliculus (SC) either directly or via the Caudate nucleus and the Substantia Nigra pars reticulata (SNpr). From SC information is forwarded to the oculomotor nerve (OMN) which drives the following saccade. Timing in the image is colour coded from early (bright) to late (dark), with estimated minimal to average latencies for each region.

Reference	Method	Primate	Region	Stimuli	Latency (ms)
(Krolak-Salmon et al., 2003)	Electrophysiology	Humans	LGN	Simple target	40-60
(Raiguel, Lagae, Guly s, & Orban, 1989)	Electrophysiology	Monkeys	V1-MT	Simple target	80-100
(Schmolesky et al., 1998)	Electrophysiology	Monkeys	LGN, V1, V2, V4, MT, FEF	Simple target	50, 66, 82, 100, 72, 75
(Munk, Nowak, Girard, Chounlamountri, & Bullier, 1995)	Electrophysiology	Monkeys	V2	Simple target	50
(Treue & Martinez-Trujillo, 2003)	Electrophysiology	Monkeys	V4		
(Raiguel, Xiao, Marcar, & Orban, 1999)	Electrophysiology	Monkeys	MT	Simple target	90
(Pack & Born, 2001)	Electrophysiology	Monkeys	MT	Simple target	70

(Thomas & Pare, 2007)	Electrophysiology	Monkeys	LIP	Simple target	100
(Bisley & Goldberg, 2003)	Electrophysiology	Monkeys	LIP	Simple target	~ 80
(Perrett, Rolls, & Caan, 1982a)	Electrophysiology	Monkeys	IT	Faces	80-160
(Hasselmo, Rolls, & Baylis, 1989)	Electrophysiology	Monkey	IT	Faces	~ 100
(Eifuku, De Souza, Tamura, Nishijo, & Ono, 2004)	Electrophysiology	Monkey	IT	Faces	~ 120
(Naya, Yoshida, & Miyashita, 2003)	Electrophysiology	Monkey	IT	Simple target	~ 85
(Kreiman et al., 2006)	Electrophysiology, LFP	Monkeys	IT	Object categorization	100
(Oram & Perrett, 1992)	Electrophysiology	Monkeys	IT	Faces	~ 110
(Liu, Agam, Madsen, & Kreiman, 2009)	iEEG	Humans	IT	Object categorization	100
(Tovee, Rolls, Treves, & Bellis, 1993)	Electrophysiology	Monkeys	IT	Faces	~ 100
(Keysers, Xiao, Földiák, & Perrett, 2001)	Electrophysiology	Monkeys	IT	Natural scenes	~ 100
(Hung et al., 2005)	Electrophysiology	Monkey	IT	Object categorization	120
(Tamura & Tanaka, 2001)	Electrophysiology	Monkeys	IT	Faces, Objects	120
(Naya, Yoshida, & Miyashita, 2001)	Electrophysiology	Monkeys	Temporal cortex	Pair association	Over 150
(Freedman, Riesenhuber, Poggio, & Miller, 2003)	Electrophysiology	Monkeys	PFC; IT	Object categorization	170-200; 100-120
(Ledberg, Bressler, Ding, Coppola, & Nakamura, 2007)	LFP	Monkey	IT; PFC; FEF		כמה שבא לי
(Schall & Hanes, 1993)	Electrophysiology	Monkeys	FEF	Simple target	100-160
(Hanes & Schall, 1996)	Electrophysiology	Monkeys	FEF	Simple target	225-250
(K. Thompson, Hanes, Bichot, & Schall, 1996)	Electrophysiology	Monkeys	FEF	Simple target	67

(Fuji, Mushiaka, & Tanji, 1998)	Electrophysiology + Stimulation	Monkeys	FEF	Natural scenes	
(Kirchner, Barbeau, Thorpe, Regis, & Liegeois-Chauvel, 2009)	iEEG	Humans	FEF	Simple target	45-60
(BAJ Reddi, 2001)	Electrophysiology	Monkeys	FEF and PFC	Simple target	120
(Desimone & Duncan, 1995)		Monkeys			
(Bar et al., 2006)	fMRI, MEG	Humans	Cortex	Object categorization	130
(Watanabe, Lauwereyns, & Hikosaka, 2003)	Electrophysiology	Monkeys	Caudate nucleus	Simple target	200-350
(Sparks, 1978)	Electrophysiology	Monkeys	SC	Simple target	150-350
(Raybourn & Keller, 1977)	Electrophysiology + Stimulation + Eye tracking	Monkeys	SC		30 to OMN
(Paré & Munoz, 1996)	Electrophysiology	Monkeys	OMN	Simple target	150
(Sommer, 1997)	Eye-tracking	Monkeys	Inter-saccade times	Simple target	150-200
(H. Kirchner & S. Thorpe, 2006)	Eye-tracking	Humans	Eye tracking	Natural scenes	120-130
(Barbeau et al., 2008)	iEEG	Humans	Fusiform gyrus	Faces	110
(Mormann et al., 2008)	Electrophysiology	Humans		Object categorization	Over 200
(S Thorpe et al., 1996)	EEG	Humans	Full procedure	Object categorization	150
(SJ Thorpe & Fabre-Thorpe, 2001)		Monkeys	All		
(Van Essen & Maunsell, 1983)		Monkey	All		
(Bullier, 2001)		Monkey	All		

Table 6 - Studies linking timing and interaction between brain regions and attention

Data in the table is compiled from a variety of studies, taken as a conservative estimate based on the majority of latency evidence for each brain region. Brain region estimates from primates were taken as is, although the reader should assume that due to the differences in brain sizes there should be increased latencies in humans. We used either the best timing in each region, or the average where these were provided numerically, or an estimate based on the figures where no exact number was provided. “Simple target” are shapes, Gabor filters, moving dots, bar, etc., while object or faces identification, and categorization are termed “Object categorization”.

Given the short latency it seems that the SC is the more likely candidate. Additionally it would seem unlikely that the frontal eye fields (FEF) would comprise two functionally distinct mechanisms of saccade generation. Some aspects of our results also match observed neurophysiology of the SC. In monkey SC the activity of build-up neurons has been shown to be related to the probability of generating a saccade to that area, and that this change in activity is predictive of saccadic latency (BA Reddi & Carpenter, 2000). If we take a build-up neuron to represent a LATER unit then we don't see any changes in its mean rate of rise but instead we see changes in its starting value based on target expectation; changing prior probability. This would produce the swivel we see in our results.

Increasing the activity of build-up neurons, and thus the prior probability, locally, where a target is expected, has further predicted effects. It reduces the increase in activity needed to trigger a saccade, thus in race-to-threshold competition with main distribution saccades it is more likely to win, and thus there should be an increase in the number of early saccades. In competition with other build-up neurons it is more likely to win and trigger a saccade, thus of

the population of early saccades more should land on “expected” targets. Basso and Wurtz also showed that the more visual targets there are the lower the general activity of build-up neurons in monkey SC; a kind of lateral inhibition (Basso & Wurtz, 1998). Thus on the first fixation where targets are unexpected and there are multiple visual targets this inhibition produced the reverse effect and makes the generation of an ES very unlikely. This is a sensible mechanism given that the SC is itself incapable of choosing what it should look at itself.

Thus we hypothesise that the SC acts as a sort of “cache”. By the second fixation targets are predictable and thus a map could be “loaded” into SC as prior probabilities in the form of build-up neuron activity. This explains the increased proportion and accuracy of early saccades seen in our results, as well as the swivel with increasing fixations. It explains why build-up neuron activity is indicative of the probability of a saccade being generated to that area and why it is predictive of saccadic latency.



Predicting Human Gaze Using Low-Level Saliency Combined with Face Detection

*Prediction is very hard, especially
about the future.*

Niels Bohr

C¹² ommonalities between different individuals' fixation patterns allow computational models to predict where people look, and in which order (Cerf et al., 2007). There are several models for predicting observers' fixations (Dickinson et al., 1994), some of which are inspired by putative neural mechanisms. A frequently referenced model for fixation prediction is the Itti et al. saliency map model [SM] (Itti et al., 1998). This “bottom-up” approach is based on contrasts of intrinsic image features such as color, orientation, intensity, flicker, motion, and so on, without any explicit information about higher-order scene structure, semantics, context, or task-related (“top-down”) factors, which may be crucial for attentional allocation (Yarbus, 1967). Such a bottom-up saliency model works well when higher-order semantics are reflected in low-level features (as is often the case for isolated objects, and even for reasonably cluttered scenes), but tends to fail if other factors dominate: e.g., in search tasks (Henderson et al., 2007), strong contextual effects (Torralba, Oliva, Castelhana, & Henderson, 2006), or in free viewing of images without clearly isolated objects, such as forest scenes or foliage (W. Einhauser & Konig, 2003). Here, we improve the standard saliency model by adding a “face channel” based on an established face detector algorithm. Although there is an ongoing debate regarding the exact mechanisms which underlie face detection, there is no argument that a normal subject (in contrast to autistic patients) will not interpret a face purely as a reddish blob with four lines, but as a much more significant

¹² This chapter is partially based on: Cerf M, Harel J, Einhauser W, Koch C, “Predicting human gaze using low-level saliency combined with face detection”, *Advances in Neural Information Processing Systems (NIPS)*, Vol. 20, Pages 241-248, 2007.

entity (Hershler & Hochstein, 2005; VanRullen, 2006). In fact, there is mounting evidence of infants' preference for face-like patterns before they can even consciously perceive the category of faces (Simion & Shimojo, 2006), which is crucial for emotion and social processing (R Adolphs, 2002; Barton, 2003).

Face detection is a well-investigated area of machine vision. There are numerous computer-vision models for face detection with good results (Roth, Yang, & Ahuja, 2000; Sung & Poggio, 1998). One widely used model for face recognition is the Viola & Jones (Viola & Jones, 2001) feature-based template matching algorithm [VJ].

There have been previous attempts to incorporate face detection into a saliency model. However, they have either relied on biasing a color channel toward skin hue (D. Walther, 2006) — and thus being ineffective in many cases, as well as not being face-selective *per se* — or they have suffered from lack of generality (Breazeal & Scassellati, 1999). We here propose a system which combines the bottom-up saliency map model of Itti et al. (Itti et al., 1998) with the Viola & Jones face detector.

The contributions of this study are: (1) A novel saliency model which combines a face detector with intensity, color, and orientation information. (2) Quantitative results on two versions of this saliency model, including one extended from a recent graph-based approach, which shows that, compared to previous approaches, it better predicts subjects' fixations on images with faces, and predicts as well otherwise.

Methods

Combined face detection with various saliency algorithms

We tried to predict the attentional allocation via fixation patterns of the subjects using various saliency maps. In particular, we computed four different saliency maps for each of the images in our data set: (1) a saliency map based on the model of (Itti et al., 1998) (SM), (2) a graph-based saliency map according to (Harel, Koch, & Perona, 2007) (GBSM), (3) a map which combines SM with face-detection via VJ (SM+VJ), and (4) a saliency map combining the outputs of GBSM and VJ (GBSM+VJ). Each saliency map was represented as a positive valued heat map over the image plane. SM is based on computing feature maps, followed by center-surround operations which highlight local gradients, followed by a normalization step prior to combining the feature channels. We used the “Maxnorm” normalization scheme which is a spatial competition mechanism based on the squared ratio of global maximum over average local maximum. This promotes feature maps with one conspicuous location to the detriment of maps presenting numerous conspicuous locations. The graph-based saliency map model (GBSM) employs spectral techniques *in lieu* of center surround subtraction and “Maxnorm” normalization, using only local computations. GBSM has shown more robust correlation with human fixation data compared with standard SM (Harel et al., 2007).

For face detection, we used the *Intel Open Source Computer Vision Library* “OpenCV” (Bradski, Kaehler, & Pisarevsky, 2005) implementation of (Viola & Jones, 2001). This implementation rapidly processes images while achieving high detection rates. An efficient

classifier built using the Ada-Boost learning algorithm (Freund et al., 1999) is used to select a small number of critical visual features from a large set of potential candidates. Combining classifiers in a cascade allows background regions of the image to be quickly discarded, so that more cycles process promising face-like regions using a template matching scheme.

The detection is done by applying a classifier to a sliding search window of 24x24 pixels. The detectors are made of three joined black and white rectangles, either up-right or rotated by 45°. The values at each point are calculated as a weighted sum of two components: the pixel sum over the black rectangles and the sum over the whole detector area. The classifiers are combined to make a boosted cascade with classifiers going from simple to more complex, each possibly rejecting the candidate window as “not a face” (Bradski et al., 2005). This implementation of the *facedetect* module was used with the standard default training set of the original model. We used it to form a “Faces conspicuity map”, or “Face channel” by convolving delta functions at the (x,y) detected facial centers with 2D Gaussians having standard deviation equal to estimated facial radius. The values of this map were normalized to a fixed range.

For both SM and GBSM, we computed the combined saliency map as the mean of the normalized color (C), orientation (O), and intensity (I) maps (Itti et al., 1998):

$$\frac{1}{3}(N(I) + N(O) + N(C))$$

And for SM+VJ and GBSM+VJ, we incorporated the normalized face conspicuity map (F) into this mean (see Figure 69):

$$\frac{1}{4}(N(I) + n(o) + n(c) + n(F))$$

This is our combined face detector/saliency model. Although we could have explored the space of combinations which would optimize predictive performance, we chose to use this simplest possible combination, since it is the least complicated to analyze, and also provides us with first intuition for further studies.



Figure 69 - Modified saliency model

An image is processed through standard (Itti et al., 1998) color, orientation, and intensity multi-scale channels, as well as through a trained template-matching face detection mechanism. Face coordinates and radius from the face detector are used to form a face conspicuity map (F), with peaks at facial centers. All four maps are normalized to the same dynamic range, and added with equal weights to a final saliency map ($SM+VJ$, or $GBSM+VJ$). This is compared to a saliency map which only uses the three bottom-up features maps (SM or $GBSM$).

Results

Assessing the saliency map models

We ran VJ on each of the 200 images used in the free viewing task, and found at least one face detection on 176 of these images, 148 of which actually contained faces (only two images with faces were missed). For each of these 176 images, we computed four saliency maps (SM, GBSM, SM + VJ, GBSM + VJ) as discussed above, and quantified the compatibility of each with our scanpath recordings, in particular fixations, using the area under an ROC curve. The ROC curves were generated by sweeping over saliency value thresholds, and treating the fraction of nonfixated pixels on a map above threshold as false alarms, and the fraction of fixated pixels above threshold as hits (Peters et al., 2005; Tatler et al., 2005). According to this ROC fixation “prediction” metric, for the example image in Figure 47, all models predict above chance (50%): SM performs worst, and GBSM + VJ best, since including the face detector substantially improves performance in both cases.

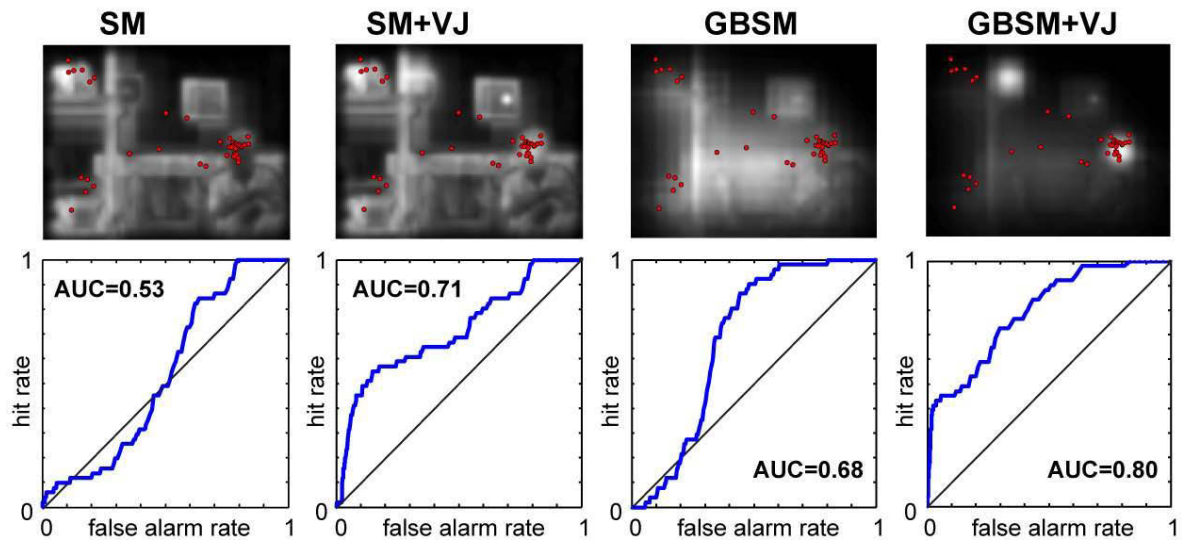


Figure 70 - Comparison of the area-under-the-curve (AUC) for an image

Top panels. Image (chosen arbitrarily) with the 49 fixations of the 7 subjects (red). First central fixations for each subject were excluded. From left to right, saliency map model of Itti et al. (SM), saliency map with the VJ face detection map (SM + VJ), the graph-based saliency map (GBSM), and the graph-based saliency map with face detection channel (GBSM + VJ). Red dots correspond to fixations.

Lower panels depict ROC curves corresponding to each map. Here, GBSM+VJ predict fixations best, as quantified by the highest AUC.

Across all 176 images, this trend prevails (Figure 71): first, all models perform better than chance, even over the 28 images without faces. The SM + VJ model performed better than the SM model for 154/176 images. The null hypothesis to get this result by chance can be rejected at $p < 10^{22}$ (using a coin-toss sign-test for which model does better, with uniform null-

hypothesis, neglecting the size of effects). Similarly, the GBSM + VJ model performed better than the GBSM model for 142/176 images, a comparably vast majority ($p < 10^{-3}$) (see Figure 71, right). For the 148/176 images with faces, SM + VJ was better than SM alone for 144/148 images ($p < 10^{-20}$), whereas VJ alone (equal to the face conspicuity map) was better than SM alone for 83/148 images, a fraction that fails to reach significance. Thus, although the face conspicuity map was surprisingly predictive on its own, fixation predictions were much better when it was combined with the full saliency model.

For the 28 images without faces, SM (better than SM + VJ for 18) and SM + VJ (better than SM for 10) did not show a significant difference, nor did GBSM versus GBSM + VJ (better on 15/28 compared to 13/28, respectively). However, in a recent follow-up study with more nonface images, we found preliminary results indicating that the mean ROC score of VJ-enhanced saliency maps is higher on such nonface images, although the median is slightly lower, i.e., performance is much improved when improved at all – indicating that VJ false positives can sometimes enhance saliency maps.

In summary, we found that adding a face detector channel improves fixation prediction in images with faces dramatically, while it does not impair prediction in images without faces, even though the face detector has false alarms in those cases.

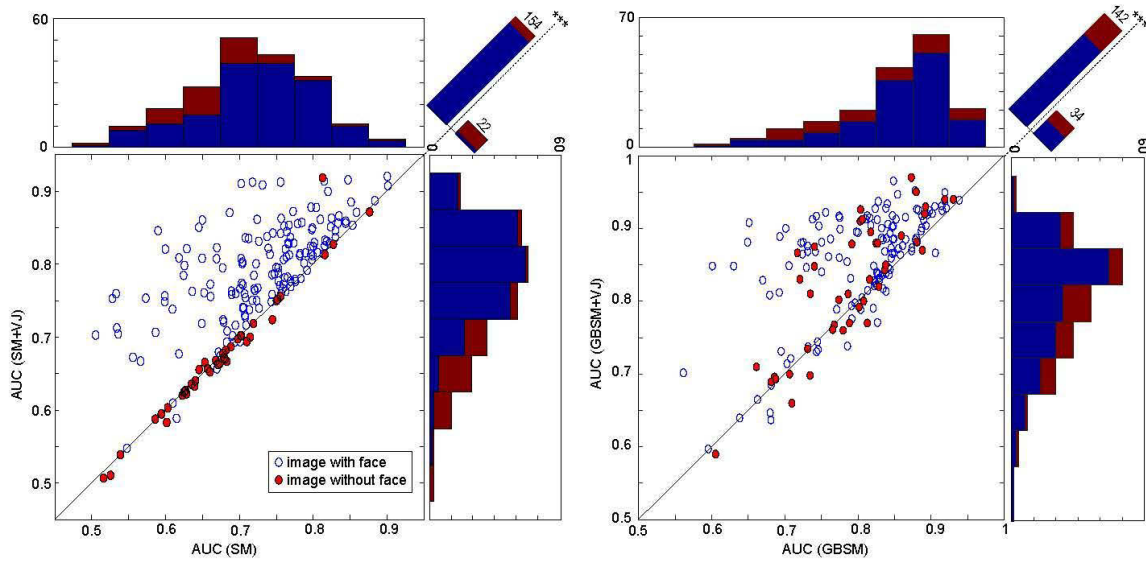


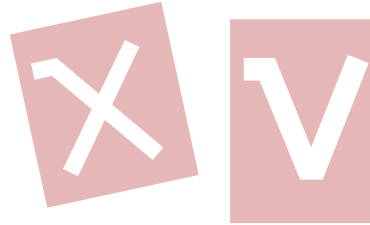
Figure 71 - Comparisons of the SM, SM + VJ and GBSM, GBSM + VJ

Scatter-plots depict the area under ROC curves (AUC) for the 176 images in which VJ found a face. Each point represents a single image. Points above the diagonal indicate better prediction of the model including face detection compared to the models without face channel. Blue markers denote images with faces; red markers images without faces, i.e., false positives of the VJ face detector). Histograms of the SM and SM + VJ (GBSM and GBSM + VJ) are depicted to the top and left (binning: 0.05); colorcode as in scatterplots.

Discussion

Since faces are fixated on within the first few fixations, independent of the task, we used this powerful trend to introduce a new saliency model, which combined the “bottom-up” feature channels of color, orientation, and intensity, with a special face-detection channel, based on the Viola & Jones algorithm. The combination was linear in nature with uniform weight distribution for maximum simplicity. In attempting to predict the fixations of human subjects, we found that this additional face channel improved the performance of both a standard and a more recent graph-based saliency model (almost all blue points in Figure 71 are above the diagonal) in images with faces. In the few images without faces, we found that the false positives represented in the face-detection channel did not significantly alter the performance of the saliency maps — although in a preliminary follow-up on a larger image pool we found that they boost mean performance. Together, these findings point towards a specialized “face channel” in our vision system, which is subject to current debate in the attention literature (VanRullen, 2006).

In conclusion, inspired by biological understanding of human attentional allocation to meaningful objects — faces — we presented a new model for computing an improved saliency map which is more consistent with gaze deployment in natural images containing faces than previously studied models, even though the face detector was trained on standard sets. This suggests that faces always attract attention and gaze, relatively independent of the task. They should therefore be considered as part of the bottom-up saliency pathway.



Computer Model of Faces and Text Gaze Attraction

These economic downturns are very difficult to model and predict, but sophisticated econometric modeling firms, like Chase Econometrics and A.I.G have successfully predicted 14 of the last 3 recessions.

Source unknown

S¹³imilar to the way by which we analyzed the ability of a computer model to predict the fixations of subjects free-viewing images with faces in the previous chapter, we tried to test how a computer model will do with other high-level cues such as text and faces. This study addresses the question of whether any high-level object can be used for improved prediction, or only ones that have more meaningful properties.

¹³ This chapter is partially based on: Cerf M, Frady P, Koch C, "Faces and text attract gaze independent of the task: Experimental data and computer model", *Journal of Vision*, 2009

Methods

The scanpaths of the subjects in the free viewing task were used to validate the predictions of subjects' attention allocation by a computer model.

Prior work shows that a biologically inspired bottom-up driven saliency model can predict subjects' fixation to an accuracy of about 53% for images containing natural scenes (Peters et al., 2005). More so, prior work by Cerf *et. al* showed that inclusion of a face map that is combined with the bottom-up feature maps yields better performance in predicting subjects' fixations (Cerf, Harel, Einhauser et al., 2008).

We here tested the ability to improve fixation prediction by adding a channel containing the exact location of the faces, text or cell phones in the image. We manually defined a minimally sized 'region of interest' (ROI) around each face, text, or cell phone, creating a binary heatmap describing the location of the entities. The original saliency map is computed as an average of 3 channels: Intensity, Orientation and Color (Itti et al., 1998):

$$S = \frac{1}{3}[\mathcal{N}(I) + \mathcal{N}(C) + \mathcal{N}(O)] \quad (1)$$

The modified saliency map adds the extra entity channel (E) (see Figure 72 for illustration):

$$S = \frac{1}{4}[\mathcal{N}(I) + \mathcal{N}(C) + \mathcal{N}(O) + \mathcal{N}(E)] \quad (2)$$

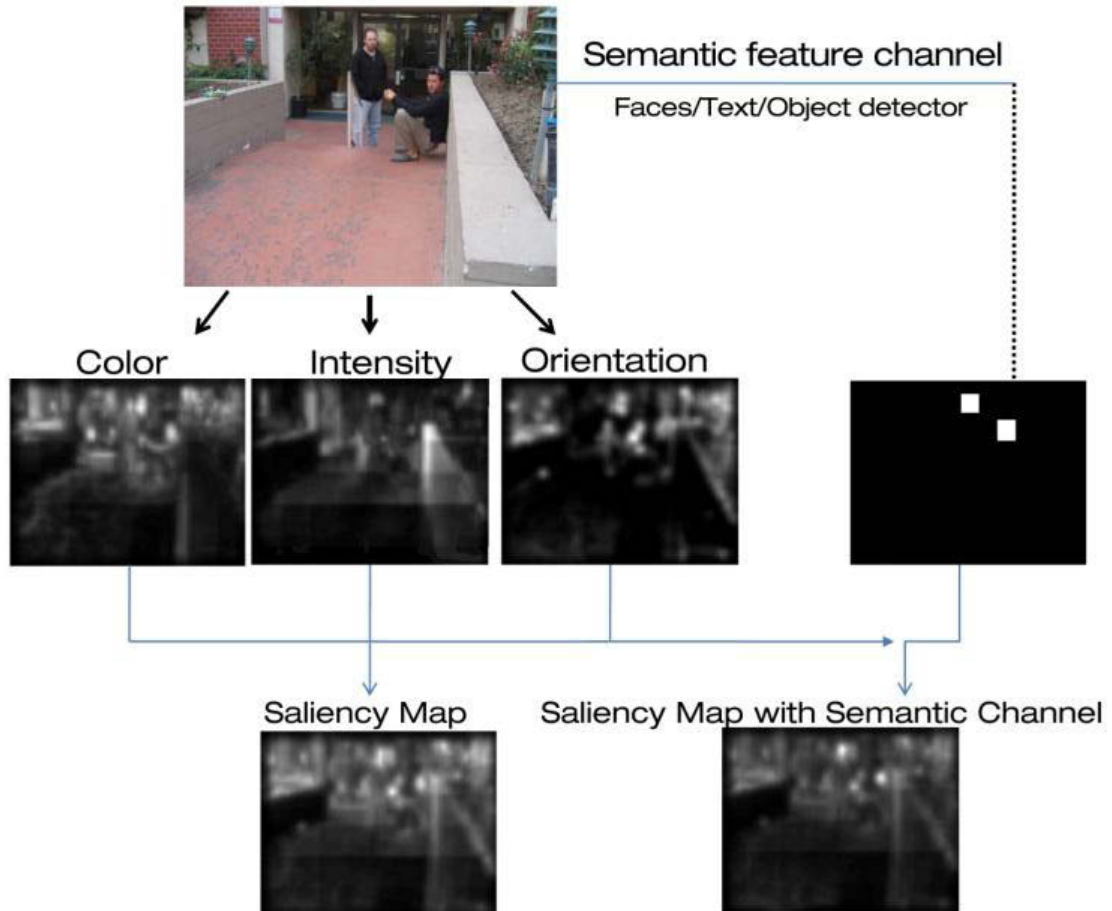


Figure 72 - Illustration of the combined saliency map with a semantic channel added

An example image is fed into feature channels for color, intensity and orientation as well as into a fourth channel for faces/text/cell phone. The combined feature channels were normalized and formed a modified saliency map that was compared to the original saliency map.

The combination was linear with uniform weight for simplicity. Performance of the saliency maps were measured by the receiver operating characteristic (ROC) curve. For each saliency map, the hit rate was calculated by determining the locations where the saliency map was above threshold and a fixation was present in these supra-threshold regions. Similarly, the false alarm rate was calculated by measuring the locations at which the saliency map was above threshold and there was no fixation present (Cerf, Harel, Huth et al., 2008). The ROC curve was generated by varying the threshold to cover all possible ranges of values the saliency map produces. The area under the ROC curve (AUC) is a general measure of how well the saliency map predicts fixations.

The AUCs were normalized by an “ideal AUC”, which measures how well the subject’s fixations predicted each other. This ideal AUC reflects an upper-bound to how well our model can predict subjects’ fixations. The AUC normalization was done such that a value of 100% would reflect the ideal AUC and a value of 50% would reflect chance. The ideal AUC was calculated by performing a similar ROC analysis on each subject using the fixation pattern of all other subjects in place of the saliency map.

This leave-one-out analysis results in an ideal AUC of $78.6\% \pm 6.1$ for faces, $78.4\% \pm 5.5$ for text, and $79.2\% \pm 6.0$ for cell phones under the free-viewing condition. None of these values are significantly different from another. This shows that the inter-subject variability is consistent

for the image sets, suggesting that there is little difference between the visual complexities of the image sets.

Results

Model analysis

In order to improve the predictive performance of the saliency algorithm, we added specific channels dedicated to faces, text and cell phones to the standard saliency map model. We calculated how well the new map predicts the locations of subjects' fixations.

The performance of the standard saliency map for viewing of images containing faces was on average 79.0% (normalized AUC). Adding the face channel increased predictability to 87.4%. Predictions of fixation location for subjects viewing text improved from 77.3% to 84.8%. Both increases in predictability are significant ($p < 10^{-6}$ for both, paired t-test). The text channel leads to improvements for every image containing text. For cell phone images, the mean AUC improved slightly but significantly (from 77.0% to 78.1%; $p < 10^{-8}$). Figure 73 compares the AUC for each of 231 individual images with the standard saliency map to the new saliency map.

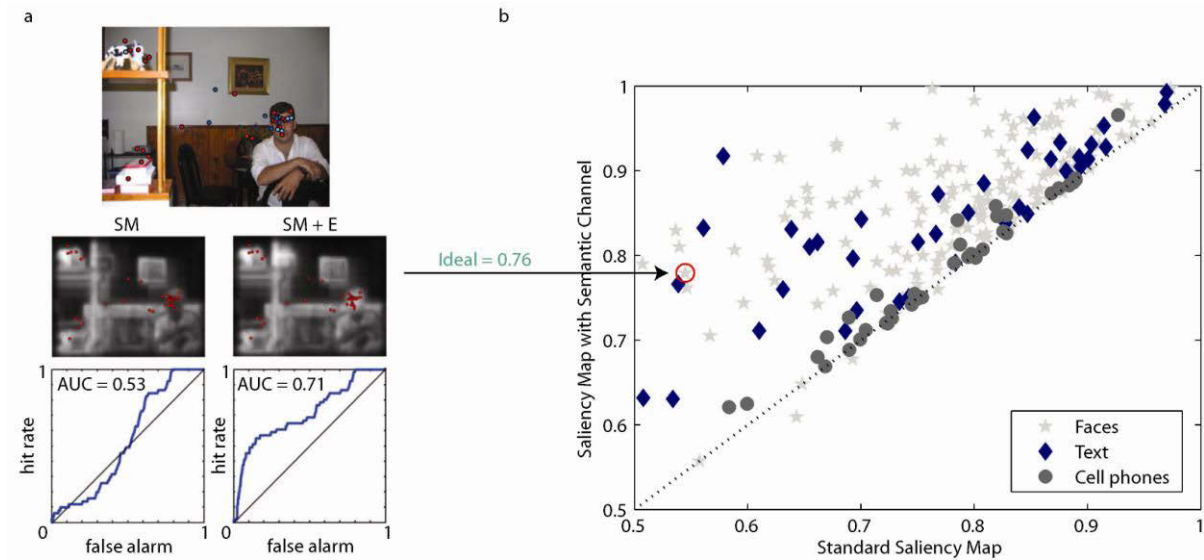


Figure 73 - Performance comparison for all 231 images

left. An example of the way by which each point in the scatter diagram was calculated. For the image pictured, the fixations of all subjects were superimposed and were compared using the ROC curve to both the standard Saliency Model (SM) and the Saliency Model with high-level entity channel (SM + E). The ROC curve is created by comparing the hit rate and false alarm rate for all possible thresholds. The AUC for each map is normalized by the ideal AUC.

right. Each symbol represents the model's performance predicting subjects' fixations on a particular image. Shapes and colors in the scatter-plot indicate the different categories. Symbols above the diagonal indicate an improvement in the saliency map model with the inclusion of a high-level channel. Images with faces or text are improved by the addition of a high-level channel. Images with faces are improved the most while images with cell phone channel inclusion show only a marginal improvement.

Discussion

In chapter 14 we showed that high-level cues, namely faces and text, that carry large amounts of semantic content act as attractors of attention. In this model we tried to use this information to predict the fixation patterns of observers viewing images with faces/text.

Implementing these attractors in a bottom-up way into the Itti & Koch saliency algorithm shows large improvements of 14% (average AUC over all faces-containing images) in predicting the scanpaths of human observers for images with faces. Same trend (6.1% increase) is shown for text. Since the algorithm is based on common low-level features that are considered to be part of attention deployment, improvements using these higher-order channels suggest that their influence is based on more than just the summation of their low-level features. Saliency models with additional high-level channels can be beneficial not only for the improvement of fixations prediction, but can also serve as a measure (using the methods we demonstrate here for comparisons of chance viewing with observer viewing) to assess the importance of a high-level feature in a scene. Importantly, since these feature-based saliency models are mainly predictive of the first fixations in a scene, it is notable that while features that indeed attract the early fixations yield an improvement in the model when added as an extra high-level channel, ones that are not looked at early will not increase the performance, yet will not necessarily hurt it. Thus, the same method can in fact be used in an opposite manner to test the relevance of the feature for our early visual attention allocation. If an additional feature channel significantly improves the predictive performance then it is regarded important to us. Additionally, regardless of a feature channel's effect on the predictive performance, it can still be added without hurting the model. In a sense, we can even take this feature's importance in acting as

an attractor to a level where it can be used for decoding of the content of an image, based on the fixation patterns of multiple observer attending to the scene (Cerf, Harel, Huth et al., 2008).

To formally evaluate the idea that faces and text are intrinsically salient, we compared predictions that the standard bottom-up saliency model makes for fixations in natural scenes. We used the predictions of the saliency algorithm enhanced by an additional map (Figure 72). The standard saliency model does not encode anything high-level or cognitive but only operates on center-surround maps defined at nine distinct scales for orientation, intensity, and color. The additional map encodes the location of all faces, text, and cell phones in the images. The saliency map analysis attempts to remove the contribution of low-level features from the salience of these entities, thus revealing only high-level features that contribute to gaze position. We see a large increase in performance when adding the face and text channels, suggesting that much of the salience of these features cannot be explained by their low-level features alone. Mainly, this can be shown by the lack of improvement for the cell phone channel, although it was, too, added to the saliency map in the same fashion.

Notice, interestingly, the comparison of the results from the “ideal AUC” to those of the standard saliency map” (Figure 70) show that the normalization by the average variability across subjects (acting as de-facto upper-bound for the performance of the saliency model) yields an improvement of about 0.1 on average.

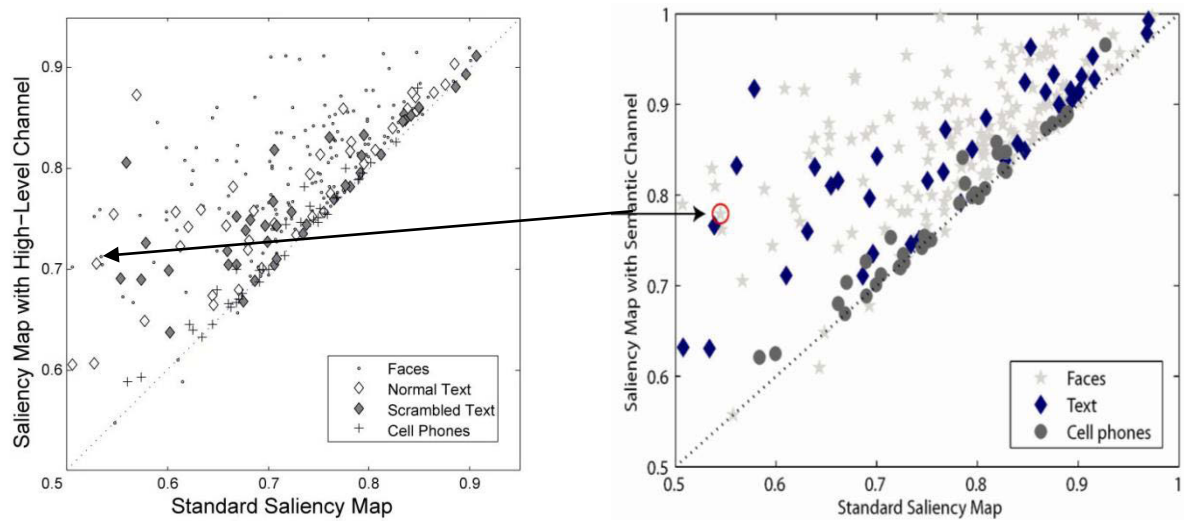


Figure 74 - Comparison of the Ideal AUC normalization and the saliency models

Left. Comparison of the standard saliency model to the one with additional High-Level Cues.

Right. The data points compared, normalized by the “Ideal AUC”. Arrows mark an example image and its AUC in either panels. The improvement for the specific image is of 0.08.

Our experiments and our modeling demonstrate that faces and text are very attractive and are difficult to ignore, even if there is a real cost associated with looking at them. It remains to be seen how the single neuron substrate of faces and text reconciles the sometimes conflicting demands of bottom-up and top-down inputs. Our improved model of saliency-driven attentional deployment is of relevance to a host of military, commercial, and consumer applications. The success of incorporating additional biologically-inspired detectors for high-level cues suggests similar attention allocation patterns to those used by the brain.



Decoding What People See from Where They Look

Predicting Visual Stimuli from Scanpaths

The bottom line is this:

The brain and the eye may have a contractual relationship, in which the brain has agreed to believe what the eye sees, and in return the eye has agreed to look for what the brain wants.

Daniel Gilbert

I¹⁴n electrophysiological studies, the ultimate validation of the relationship between physiology and behavior is the decoding of behavior from physiological data alone (Hung et al., 2005; Logothetis, Pauls, & Poggio, 1995; Perrett, Rolls, & Caan, 1982b; Quiroga, Reddy, Koch, & Fried, 2007; Sato, Kawamura, & Iwai, 1980; E. Schwartz, Desimone, Albright, & Gross, 1983; Young & Yamane, 1992). If one can determine which image an observer has seen using only the firing rate of a single neuron, one can conclude that that neuron's output is highly informative about the image set. In psychophysical studies it is common to show an observer (animal or human) a sequence of images or video while recording their eye movements using an eye-tracker. Often, such studies aim to predict subjects' scanpaths using saliency maps (Yarbus, 1967), or other techniques (Goldstein, Woods, & Peli, 2007). The predictive power of a saliency model is typically judged by computing some similarity metric between scanpaths and the saliency map generated by the model (L. Itti & C. Koch, 2001). Several similarity metrics have become de facto standards, including NSS (Peters et al., 2005) and ROC (Tatler et al., 2005). A principled way to assess the goodness of such a metric is to compare its value for scanpath-saliency map pairs which correspond to the same image and different images. If this difference is systematic, one can apply the metric to several candidate saliency maps per image, and assess which saliency map yields the highest decodability.

¹⁴ This chapter is partially based on: Cerf M, Harel J, Huth A, Koch C, "Decoding what people see from where they look: predicting visual stimuli from scanpaths", Lecture Notes in Artificial Intelligence, LNAI 5395, pp. 15-26. L. Paletta and J.K. Tsotsos (Eds.) Springer, 2009

This decodability represents a new measure of saliency map efficacy. It is complementary to the current approaches: rather than predicting fixations from image statistics, it predicts image content from fixation statistics. The fundamental advantage of rating saliency maps in this way is that the score reflects not only how similar the scanpath is to the map, but also how **dissimilar it is from the maps of other images**. Without that comparison, it is possible to artificially inflate similarity metrics using saliency heuristics which increase the correlation with all scanpaths, rather than only those recorded on the corresponding image. Thus, we propose this as an alternative to the present measures of saliency maps' predictive power, and test this on established eye-tracking datasets.

The contributions of this study are:

1. A novel method for quantifying the goodness of an attention-prediction model based on the stimuli presented and the behavior.
2. Quantitative results using this method that rank the importance of feature maps based on their contribution to the prediction.

Methods

Experimental setup

In order to test if scanpaths could be used to predict which image from a set was being observed at the time it was recorded, we collected a large dataset of images and scanpaths from various earlier experiments (from the database of (Cerf et al., 2007)). In all of these previous experiments, images were presented to subjects for 2s, after which they were instructed to answer “How interesting was the image?” on a scale of 1-9 (9 being the most interesting). Subjects were not instructed to look at anything in particular; their only task was to rate the entire image. Subjects were always naïve to the purpose of the experiments. The subset of images was presented for each subject in random order.

Scenes were indoor and outdoor still images (see examples in Figure 75), containing faces and objects. Faces were in various skin colors and age groups, and exhibited neutral expressions. The images were specifically composed so that the faces and objects appeared in a variety of locations but never in the center of the image, as this was the location of the starting fixation on each image. Faces and objects varied in size. The average size was $5\% \pm 1\%$ (mean \pm s.t.d.) of the entire image – between 1° to 5° of the visual field. The number of faces in the images was varied between 1-6, with a mean of 1.1 ± 0.48 (s.t.d.). 441 images (1024 x 768 pixels) were used in these experiments altogether. Of these, 291 images were unique. The remaining 150 stimuli consisted of 50 different images that were repeated twice, but treated uniquely as they were recorded under different experimental conditions. Of the unique images, some were very similar to each other, as only foreground objects but not the background was changed. Since

we only counted finding the exact same instance (i.e., 1 out of 441) as correct prediction, in at least $\frac{150}{441} \times \frac{2}{440} = 0.15\%$ of cases a nearly correct prediction (same or very similar image) would be counted as incorrect. Hence, our datasets are challenging and the estimates of correct prediction conservative.

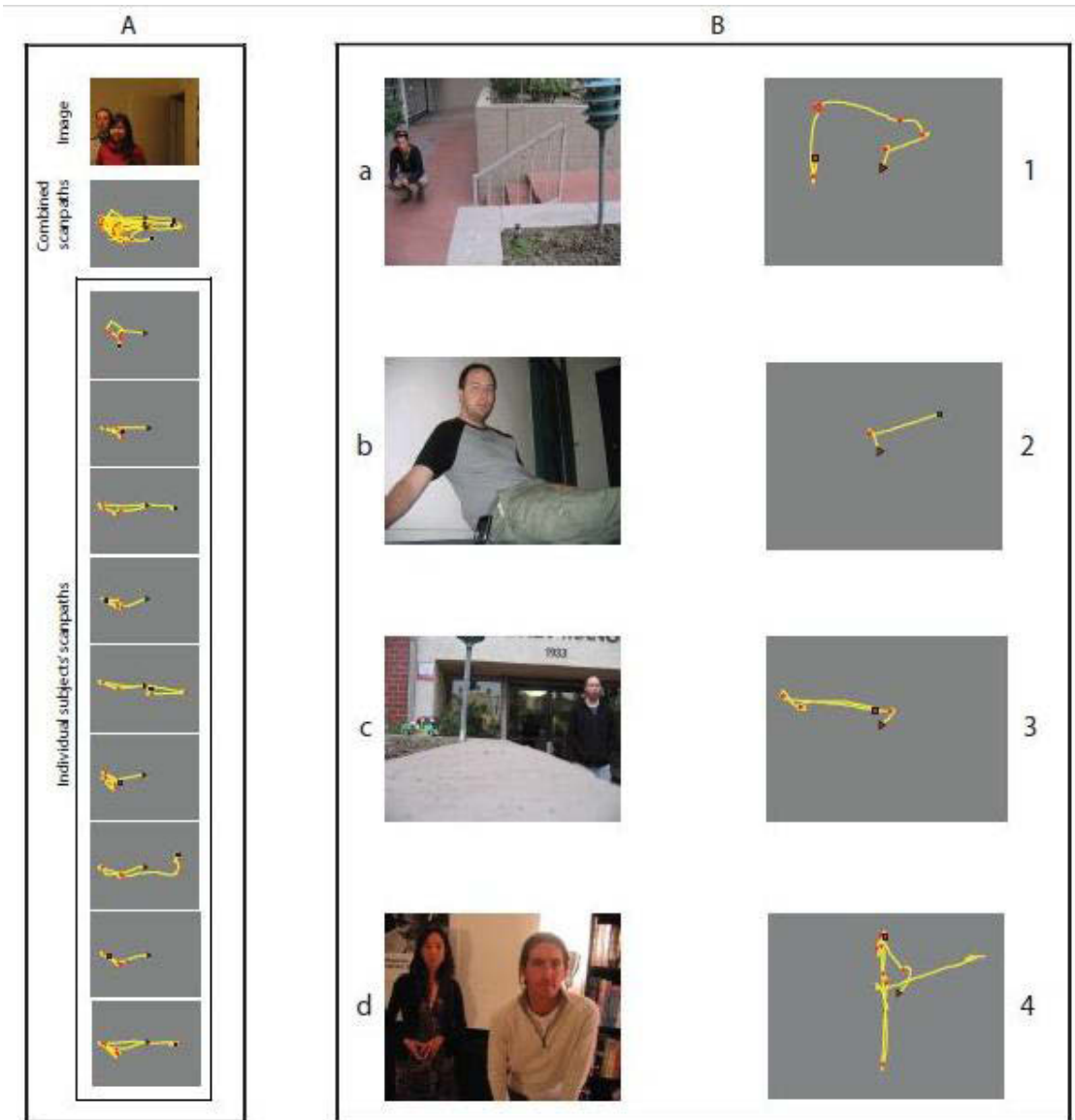


Figure 75 - Examples of scanpath/stimuli used in the experiment

A. Scanpaths of the 9 individual subjects used in the analysis for a given image. The combined fixations of all subjects were used for further analysis of the agreement across all subjects, and for analysis of the ideal subjects' pool size for decoding. The red triangle marks the first and the red square the last fixation, the yellow line the scanpath, and the red circles the subsequent fixations. **Top:** the image viewed by subjects to generate these scanpaths. The trend of visiting the faces — a highly attractive feature — yields greater decoding performance.

B. Four example images from the dataset (left) and their corresponding scanpaths for different arbitrary chosen individuals (right). Order is shuffled. See if you can match (“decode”) the scanpath to its corresponding images. (*The correct answers are: a3, b4, c2, and d1.*)

Eye-position data were acquired at 1000 Hz using an Eyelink-1000 (SR Research, Osgoode, Canada) eye-tracking device. The images were presented on a CRT2 screen (120 Hz), using Matlab's Psychophysics and eyelink toolbox extensions. Stimulus luminance was linear in pixel values. The distance between the screen and the subject was 80 cm, giving a total visual angle for each image of 28° x 21°. Subjects used a chin-rest to stabilize their head. Data were acquired from the right eye alone. Data from a total of nine subjects, each with normal or corrected-to-normal vision, were used. We discard the first fixation from each scanpath to avoid adding trivial information from the initial center fixation. Thus, we worked with $441 \times 9 = 3969$ total scanpaths.

Decoding metric

For each image, we created six different “feature maps”. Four of the maps were generated using the Itti and Koch saliency map model (Itti et al., 1998): (1) combined color-intensity-orientation (CIO) map, (2) color alone (C), (3) intensity alone (I), and (4) orientation alone

(O). A “faces” map was generated using the Viola and Jones face recognition algorithm (Viola & Jones, 2001). The sixth map, which we call “CIO+F” was a combination of the face map and the CIO map from the Itti and Koch saliency model, which has been shown to be more predictive of observers fixations than CIO (Cerf, Harel, Huth et al., 2008). Each feature map was represented as a positive valued heat map over the image plane, and downsampled substantially, in line with (L Itti & C Koch, 2001), in our case to nine by twelve pixels, each pixel corresponding to roughly 2 x 2 degrees of visual angle. Subject fixation data was binned into an array of the same size. The saliency maps and fixation data were compared using an ROC-based method (Tatler et al., 2005). This method compares saliency at fixated and nonfixated locations (see Figure 76 for an illustration of the method). We assume some threshold saliency level above which locations on the saliency map are considered to be predictions of fixation. If there is a fixation at such a location, we consider it a hit, or true positive. If there is no fixation, it is considered a false positive. We record the true positive and false positive rates as we vary the threshold level from the minimum to the maximum value of the saliency map. Plotting false positive versus true positive results in a Receiver Operator Characteristics (“ROC”) curve. We integrate the area under this ROC curve (“AUC”) to get a scalar similarity measure (AUC of 1 indicates all fixations fall on salient locations, and AUC of 0.5 is chance level). The AUC for the correct scanpath-image pair was ranked against other scanpath-image pairs (from 1 to 31 decoy images, chosen randomly from the remaining 440 to 410 images), and the decoding was considered successful only if the correct image was ranked one. In the largest image set size we tried, if any of the other 31 AUCs for scanpath/images was higher than the one of the correct match, we considered the prediction a miss (e.g., for one decoding trial the algorithm would be as follows:

1. Randomly select a scanpath out of the 3969 scanpaths.
2. Consider the image it belongs to, together with 1 to 31 randomly selected decoys. We will attempt to match the scanpath to its associated image out of this set of candidates.
3. Compute a feature map for each image in the candidate set.
4. Compute the AUC of the scanpath for each of the 2-32 saliency maps.
5. Decoding is considered successful iff the image on which the scanpath was actually recorded has the highest AUC score).

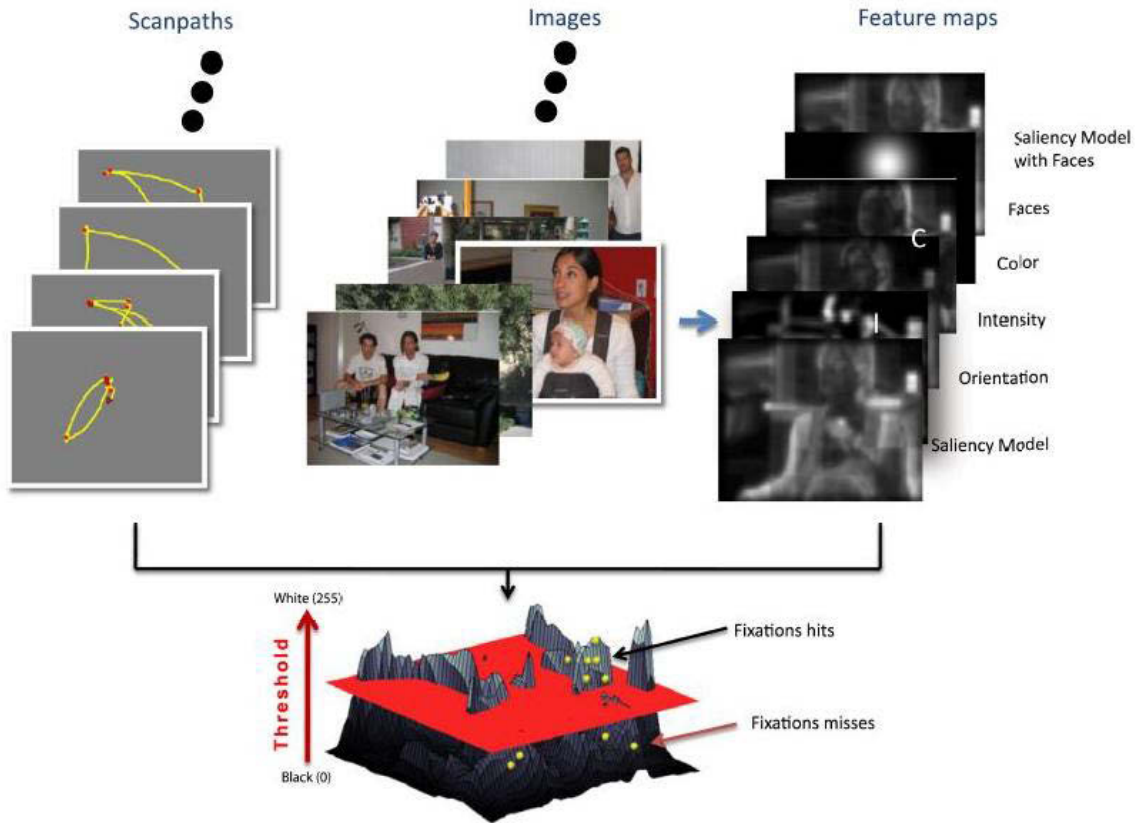


Figure 76 - Illustration of the AUC calculation

For each scanpath, we chose the corresponding image and 1-31 decoys. For each image we calculate each of the 6 feature maps (C, I, O, F, CIO, CIO+F). For a given scanpath and a feature map we then calculate the ROC by varying a threshold over the feature plane and counting how many fixations fall above/below the threshold. The area under the ROC curve (AUC) serves as a measure of agreement between the scanpath and the feature map. We then rank the images by their AUC scores, and consider the decoding correct if the highest AUC is that of the correct image.

Results

We calculated the average success rate of prediction trials, each of which consists of (1) fixations pooled over 9 subjects' scanpaths, and (2) an image set of particular cardinality, from 2 to 32, ranked according to the ROC-fixation score on one of three possible feature maps: CIO, CIO+F, or F. We used the face channel although it carries some false identifications of faces, and some misses, as it has been shown to have higher predictive power, involving high-level (semantic) saliency content with bottom-up driven features (Cerf, Harel, Einhauser et al., 2008). We reasoned that using the face channel alone in this discriminability experiment would provide a novel method of comparing it to saliency maps' predictive power.

For one decoy per image set (image set size = two), we find that the face feature map (F) was used to correctly predict the image seen by the subjects in 69% of the trials ($p < 10^{-15}$, sign test¹⁵) while the CIO+F feature map was correct in 68% ($p < 10^{-14}$), and CIO in 66% ($p < 10^{-13}$) of trials. This $F > CIO+F > CIO$ trend persists through all image set sizes. Pooling prediction trials over all image set sizes (6 sizes x 441 trials per size = 2646 trials), we find that using the F map yields a prediction that is at least as accurate as the CIO map in 89.9% of trials, with significance $p < 10^{-8}$ using the sign-test. Similarly, F is at least as predictive as CIO+F in 90.3% of trials ($p < 10^{-13}$), and CIO+F is at least as predictive as CIO in 97.8% of trials ($p < 10^{-21}$). All data points in Figure

¹⁵ The sign-test tests against the null hypothesis that the distribution of correct decodings is drawn from a binary distribution (50% for the choice of 1 of 2 images, 33% in the case of 1 of 3 images, and so forth up to 3% in the case of 1 out of 32 images). This is the most conservative estimate; additional assumptions on the distribution would yield lower p-values.

77 are significantly above their corresponding chance levels, with the least significant point corresponding to predictions using CIO with image set size 4: this results in correct decoding in 33.6% of trials, compared to 25% for chance, with null hypothesis that predictions are 25% correct being rejected at $p < 10^{-4}$.

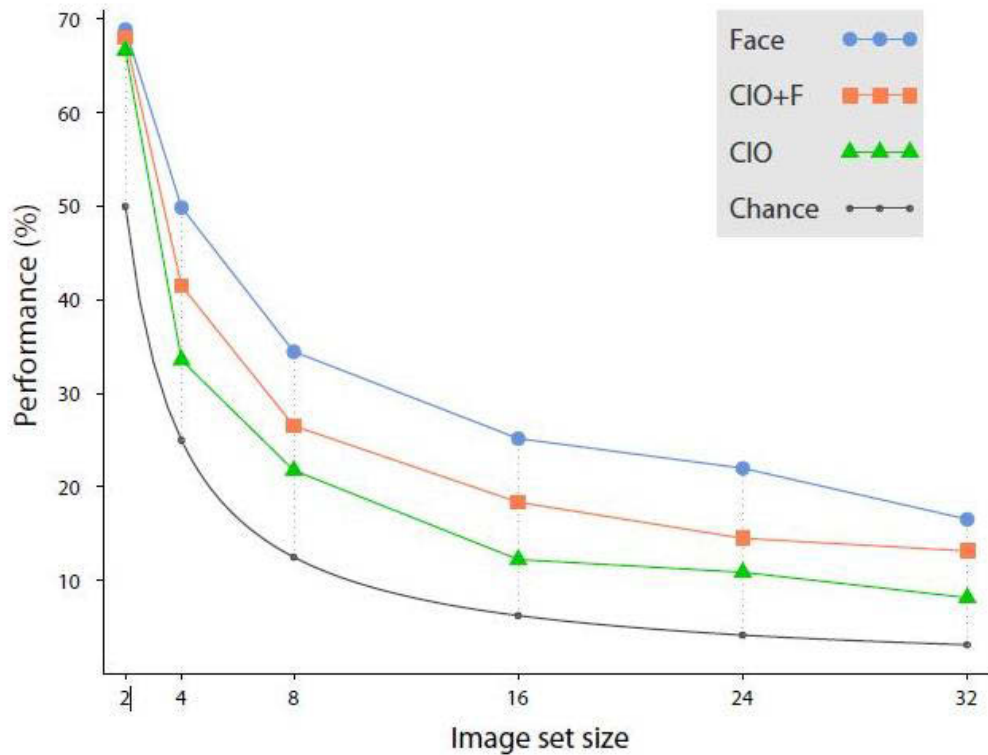


Figure 77 - Decoding performance with respect to image pool size

Decoding with scanpaths pooled over 9 subjects, we varied the number of decoy images used between 1 and 31. The larger the image set size, the more difficult the decoding. For each image set size and scanpath we calculated the ROC over 3 feature maps: a face-channel which is the output of the Viola and Jones face-detection algorithm with the given image (F), a saliency map based on the color, orientation and intensity maps (CIO), and a saliency map

combining the face-channel and the color, orientation and intensity maps (CIO+F). While all feature maps yielded a similar decoding performance for the smaller pool size, the performance was least degraded for the F map. The face feature map is higher than the CIO+F map and the two are higher than the CIO map. All maps predict above chance level – shown in the bottom line as the multiplicative inverse of the image set size.

We also tested the prediction rates when fixations were pooled over progressively fewer subjects, instead of only nine as above. For this, we used only the CIO+F map (although the face channel shows the highest decoding performance, we wanted to use a feature map that combines bottom-up features to match common attention prediction methods), and binary image trials (one decoy). One might imagine that pooling over fixation recordings from different subjects would increase the signal-to-noise ratio, but in fact we find that prediction performance only decreases (Figure 78) with more subjects. There are several possible explanations for this decrease. First, in computing the AUC, we record a correct detection (“hit”) whenever a superthreshold saliency map cell overlaps with at least one fixation, but discard information about multiple fixations at that location (i.e., a cell is either occupied by a fixation or not). Thus, the accuracy of the ROC AUC agreement between a saliency map and the fixations of multiple observers degrades with overlapping fixations. As the number of overlapping fixations increases with observers, the reliability of our decoding measure decreases. Indeed, other measures taking into account this phenomenon then can outperform the present metric. Second, if different observers exhibit distinct feature preferences (say, some prefer “color”, some prefer “orientation”, etc.), the variability in the locations of such features across an image set would contribute to the prediction in this set. It is possible that an image set

is more varied along the preferences of any one observer on average than along the pooled preferences of multiple observers. This would make it more difficult to decode from aggregate fixation sets.

The mean percentage of correct decoding for a single subject was 79% (chance is 50%), ($p < 10^{-288}$, sign test). For all combinations of 1 to 9 subjects used, the prediction was above chance (with p values below $p < 10^{16}$). The lowest prediction performance results from pooling over all nine subjects, with 66% hit rate (still significantly above chance at 50%). Figure 78 shows the prediction for each of the 9 subjects with the CIO+F feature map.

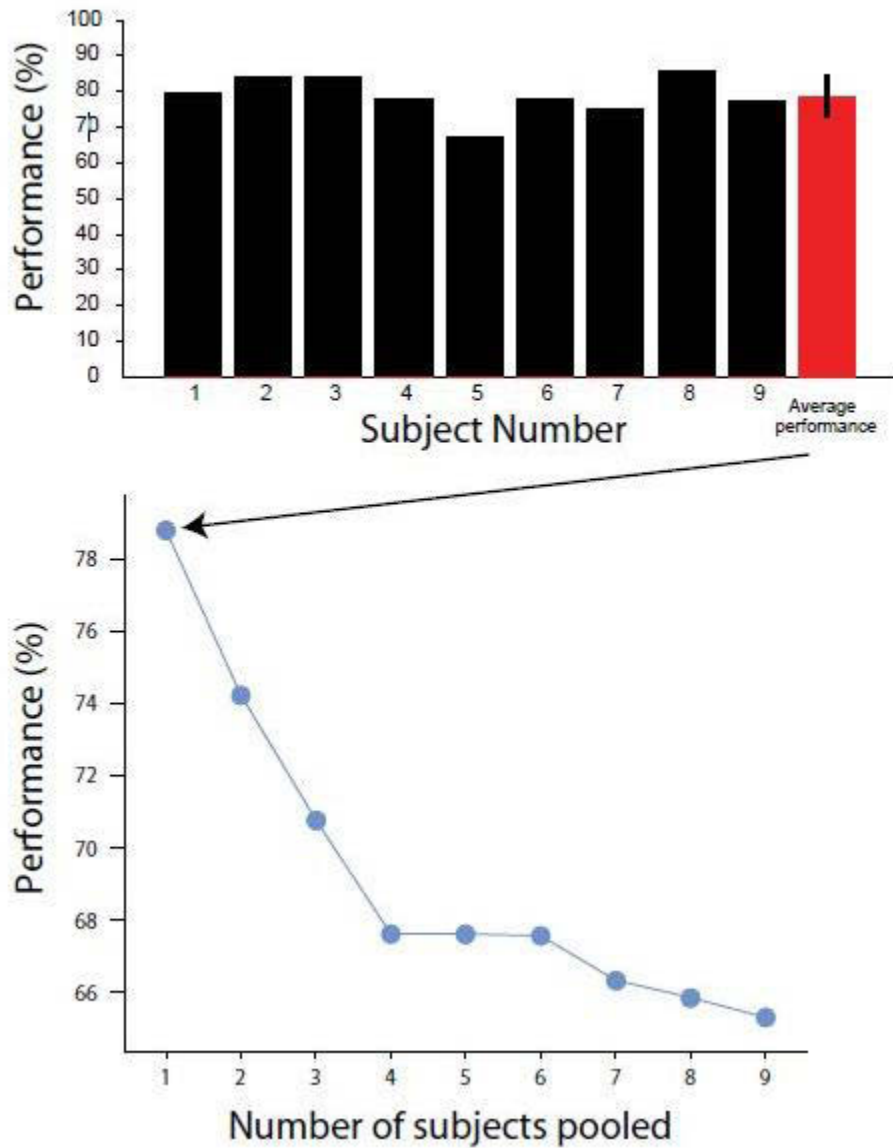


Figure 78 - Performance of the 9 individual subjects

Upper panel. For the 441 scanpaths/images, we computed the decoding performance of each individual subject. Bars indicate the performance of each subject. Red bar on the right indicates the average performance of all 9 subjects, with standard error bar. Average subject performance was 79%, with the lowest decoding performance at 67% (subject 4), and the

highest at 86% (subject 8). All values are significantly above chance (50%), with p values (sign test) below 10^{-10} .

Lower panel. Performance of various combinations of the 9 subjects. Scanpaths of 1, 2, ... 9 subjects used to determine the performance differences by using average scanpaths of multiple subjects. The performance of individual subjects shown on the leftmost point is the average of each subjects' performance as shown in the upper panel. The rightmost point is the performance of all subjects combined. Each subject pool was combined from a random choice of subjects out of the 9, reaching the pool size.

Finally, in order to test the relative contribution of each feature map to the decoding, we used our new decoding correctness rate to compare feature map types, from most discriminating to least. This was done by comparing separately each of the 6 features maps' average decoding performance for binary trials with 9 individual subjects' scanpaths. The results (Figure 79) show that out of the 6 feature maps the face channel has the highest performance (decoding performance of 82%, $p = 0$) (as shown also in Figure 77), and the intensity map has the lowest performance (decoding performance: 65%, $p < 10^{-10}$, sign test). All values are significantly above chance (50%).

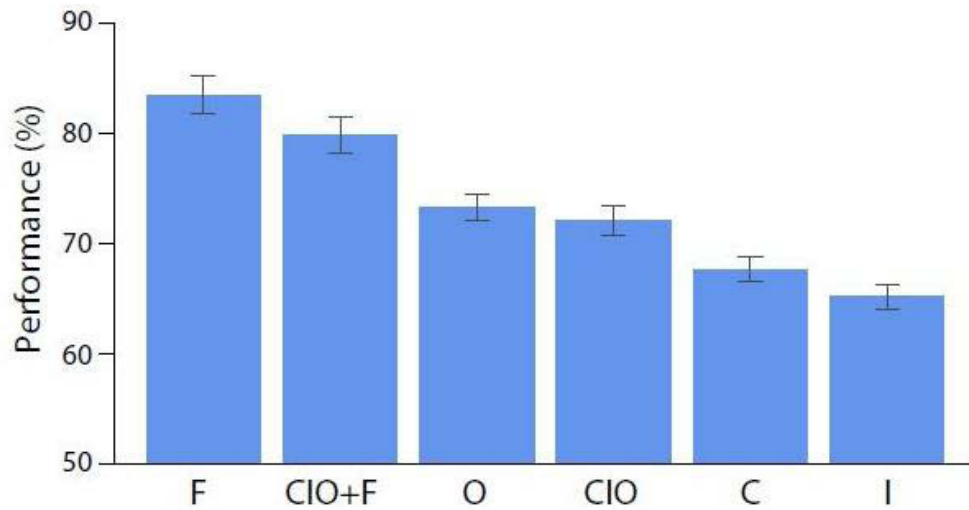


Figure 79 - Decoding performance based on feature maps used

We show the average decoding performance on binary trials using each of the 6 different feature maps, and in each trial, the scanpath of only one individual subject. Thus, for instance, the performance of the CIO + F map is exactly that shown in the average bar in Figure 78. The higher the performance the more useful the feature is in the decoding. The face channel is the most important one for this dataset.

Discussion

In this study, we investigated if scanpath data could be used to decode which image an observer was viewing given only the scanpath and saliency maps. The results were quite strong: in an experiment with 441 trials, each consisting of 32 images with scanpath data belonging to one unknown image in the set, in 73 trials (17%) the correct image was selected, a fraction much higher than chance ($\frac{1}{32}=3\%$). This leads us to propose a new metric for quantifying the efficacy of saliency maps based on image discriminability. For decoding we used the standard area under ROC curve measure with the fixations from 1 to 9 subjects on a feature map generated by popular models for fixations and attention predictions.

The “decodability” of a dataset is a score given to the combined scanpath/stimuli data for a given feature and as such can be used in various ways: we here used the decodability in order to compare ideal combined subjects' scanpath pool and feature maps' predictive power. Furthermore, we can imagine the same method being used to cluster subjects according to features that pertain specifically to them for a given dataset (i.e., if a particular set of subjects tends to look more often on an area in the images than other (Buswell, 1935), or tends to fixate on a certain object/target more (Barton, 2003), this would result in a higher decoding performance for that feature map), or as a measure of the relative amount of stimuli needed to reach a certain level of decoding performance. Our data suggests that clustering by such features to segregate between autistic and normal subjects is perhaps possible based on

differences in their looking at faces/objects (Klin, Jones, Schultz, Volkmar, & Cohen, 2002).

However, our autism subject's fixations dataset is too small to reach significance.

In line with earlier results, ours show that saliency maps using bottom-up features such as color, orientation, and intensity are relatively accurate predictors of fixation (Einhäuser et al., 2006; V Navalpakkam & Itti, 2007; Tatler et al., 2005) with a performance above 70% (Figure 79, similar to the estimate in (Peters et al., 2005). Adding the information from a face detector boosts performance to over 80%, similar to the estimate in (Cerf, Harel, Einhauser et al., 2008). It is possible that incorporating more complex, higher-level feature maps (Kayser, Nielsen, & Logothetis, 2006) could further improve performance.

Some of the images we used were very similar to each other, and so the image set could be considered challenging. Using this novel decoding metric on larger, more diverse datasets could yield more striking distinctions between the feature maps and their relative contributions to attentional allocation.

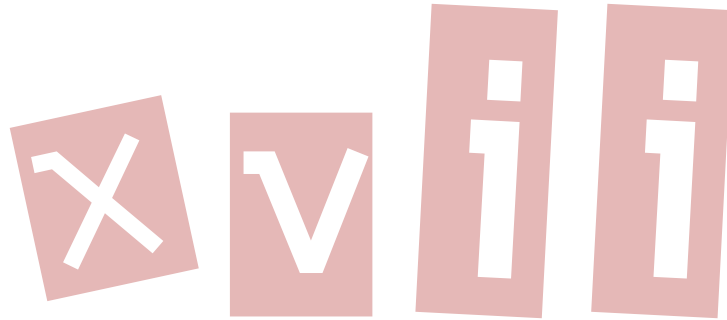
Notice that in the results, in particular in Figure 77, we computed average predictive performance using fixations pooled over all 9 scanpaths recorded per image. However, as we have shown that individual subjects' fixations are more predictive due to variability issues, these results should be even stronger than those we have included above.

A possibility for subsequent work is the prediction not of particular images from a set, but of image content. For example, is it possible to predict whether or not an image contains a face, text, or other specific semantic content based only on the scanpaths of subjects? The same kinds of stereotypical patterns we used to predict images would be useful in this kind of experiment.

Finally, one can think of more sophisticated algorithms for predicting scanpath/image pairs. For instance, one could use information about previously decoded images for future iterations (perhaps by eliminating already decoded images from the pool, making harder decoding more feasible), or a softer ranking algorithm (here we considered decoding correct only if the corresponding scanpath was ranked the highest among 32 images; one could, however, compute statistics from a soft “confusion matrix” containing all rankings so as to reduce the noise from spuriously high similarity pairs).

We demonstrated a novel method for estimating the similarity between a given set of scanpaths and images by measuring how well scanpaths could decode the images that corresponded to them. Our decoder ranked images according to saliency map/fixation similarity, yielding the most similar image as its prediction. While our decoder already yields high performance, there are more sophisticated distance measures that might be more accurate, such as ones used in electrophysiology (Quiroga et al., 2007).

Rating a saliency map relative to a scanpath based on its usability as a decoder for the input stimulus represents a robust new measure of saliency map efficacy, as it incorporates information about how dissimilar a map is from those computed on other images. This novel method can also be used for assessing images sets, for measuring the performance and attention allocation for a given set, for comparing existing saliency map performance measures, and as a metric for the evaluation of eye-tracking data against other psychophysical data.



Visual Attention Allocation in Individual with Autism

If you describe a landscape, or a cityscape, or a seascape, always make sure to put a human figure somewhere in the scene.

Why? Readers are human beings, mostly interested in human beings.

Kurt Vonnegut
To his creative writing students



fter finishing my study showing that indeed faces draw attention rapidly, independent of the task and much more than any other cue in the scene, and suggesting that this might rise from an innate face detection mechanisms in our brains I felt that my theory is solid and sound. A meeting between Christof and Ralph Adolphs during a Sunday morning run across the San Gabriel mountains proved otherwise. If you really want to show that faces attention allocation is a solid and innate aspect of humans attention allocation, a perfect way to test this would be to show that the effect doesn't hold with a population with an absence of social interest in them. People that lack that brain mechanism drawing them to social cues. If you show that indeed, these people behave differently when performing in the experiment then the results will indeed prove to be stronger and even more solid. People with Autism will be the perfect control group: they don't care about faces, or any social cues for that matter. How will they perform in this task?

Christof, I feel, was a bit reluctant to venture into this new field of research at the time. Starting over a study that already took us quite some time and with a completely new population that we had never worked with sounded like a decision that should not be made rashly. He felt that we needed to think about it thoroughly before we journeyed into this new research. Luckily an autistic subject was coming to Caltech that Thursday. We decided to try and just run one trial with this subject and see what happens. Then we will rethink the plan.

Subject DM was a very unique autism subject, with a very distinct attention allocation patterns. After running our experiment with him it was clear that there was no turning back.

Introduction

Humans are a social entity. When encountered with a scene observers commonly attend to the most socially rewarding aspect of it (R Adolphs, 1999). Prior studies suggest that this attention allocation is done using a social saliency mechanism in our brain that shifts our attentional spotlight to the most engaging resource (Amaral, 2002). However, individuals with autism are said to be less inclined to look at these social cues. Autism subjects are, thus, less likely than controls to attend to entities with higher social saliency in a scene (Klin et al., 2002). This is suggested to be either due to deficits in the mechanisms governing attention allocation, or due to differences in the weighting allotted to the social cues (Sasson et al., 2007).

One very important aspect of a social scene is the existence of other humans in it, and most importantly humans' faces (Williams, Goldstein, & Minshew, 2005). Faces usually contain very relevant social information on a scene as they carry highly rewarding social information that can assist in judging the emotional content of the scene.

Prior studies have shown that when faced with a large set of images containing faces, observers are very likely to fixate on the faces region in their very early fixations (by their second fixation 90.07% of the time, see previous chapters). This is true even if the faces take a very small portion of the frame, suggesting that indeed we are driven to look at faces early when encountered with scene containing them.

Similar studies with entities other than faces — such as text or cell phones — have shown to not have the same effect. While observers look at text early, for instance, they do so later in terms of time, and not necessarily in their first fixations (they will attend the text 90.07% of the time

by their fourth fixation, rather than their second). This effect is even stronger when the content has very little social value (cell phone or other objects), in which case they are likely only to be fixated on by the seventh or even later fixation, and are very likely to not be fixated on at all if the image viewing is limited to 2 seconds.

This tendency to look at faces early cannot be explained by saliency model based on images features (color, intensity, or orientations of the borders) alone, but have been shown to be dependant mostly on knowing where the face in the scene is, suggesting a bottom-up mechanism in our brains that identifies the face locations in the scene early, shifting our attention to that area in the image. If individuals with autism indeed lack the social need to look at faces, but are rather interested in other various aspects of the scene, then there should be a complementary feature map to that of faces for normals that can account for what attract autism subjects' attention. This is true only if the degree of variance among subjects with autism is as low as that of normal subjects upon looking at a fixed set of images.

In this study we test the degree by which subjects with autism tend to look at faces in a scene. We compare the results in two different tasks, one which allows the subjects to look at the scene somewhat freely without any instructions to attend to something specific, and one where subjects are instructed to carefully attend to a given entity in the scene. We compare the results of autism subjects to those of normals when looking at scenes with human faces and investigate the magnitude of differences in fixations to these relevant social cues between the two groups.

In order to validate our results and suggest something comparable to the attention allocation method utilized by individuals with autism when viewing images with social content, we compare the results to the viewing of images with text by the controls. Under the hypothesis

that autism attention allotted to social scenes is governed partially by the training to look at faces, rather than an innate tendency to attend to those, we have subjects view sequence of images with text (which in our case are regarded as a cue that is very relevant but is likely to not have an innate mechanism in our brain, since text is an evolutionarily recent development, and very different across cultures) and compare the viewing patterns of autism subjects looking at faces to that of controls looking at text.

Methods

15 normal subjects (ages: 18–47, 27.47 ± 7.62 years old; 11 males; IQ: 115.71 ± 8.54) and 10 subjects with autism (ages: 18–45, 29.5 ± 9.96 ; see Table 5 below for details on subjects) viewed a set of 250 images (1024x768) in a three-phases experiment. 200 of the images included head-on faces of various people; Of the 50 remaining images 25 contained no faces but were otherwise identical, allowing a comparison of viewing a particular scene with and without a face. In the first (“free-viewing”) phase of the experiment, 200 of these images (the same subset for each subject) were presented to subjects for 2 seconds, after which they were instructed to answer “How interesting was the image?” using a scale of 1–9 (9 being the most interesting). Subjects were not instructed to look at anything in particular; their only task was to rate the entire image. In the second (“search”) phase, subjects viewed another 200 image subset in the same setup, only this time they were initially presented with a probe images (either a face, or an object in the scene: banana, cell phone, toy car, etc.) for 600 ms after which one of the 200-image appeared for 2 seconds. They were then asked to indicate whether that image contained the probe. Half of the trials had the target probe present. In half of those the probe was a face. We used the second task to test if there are any differences in the fixation orders and viewing patterns when subject subjectively choose what to attend to, versus trials where they are directed to attend to faces/objects. For the search task, subjects received acoustic feedback at the end of each trial indicating correct/incorrect response. In the third phase, subjects performed a 100-image recognition memory task where they had to answer with y/n as to whether they had seen the image before.

The images were introduced as “regular images that one can expect to find in an everyday personal photo album”. Scenes were indoor and outdoor still images (see examples in Figure 80). Images included faces in various skin colors, age groups, and positions (no image had the face at the center as this was the starting fixation location in all trials). A few images had face-like objects (like animal faces), and objects that had irregular faces in them (masks, the Egyptian sphinx face, etc.)¹⁶. The average face was $5\% \pm 1\%$ (mean \pm s.t.d.) of the entire image – between 1° to 5° of the visual field; we also varied the number of faces in the image between 1-6, with a mean of 1.1 ± 0.48 . Image order was randomized throughout, and subjects were naïve to the purpose of the experiment. Subjects fixated on a cross in the center before each image onset. Eye-position data were acquired at 1000 Hz using an Eyelink-1000 (SR Research Osgoode, Canada) eye-tracking device. The images were presented on a CRT screen (120 Hz), using Matlab’s Psychophysics and eyelink toolbox extension. Stimulus luminance was linear in pixel values. The distance between the screen and the subject was 80 cm, giving a total visual angle for each image of $28^\circ \times 21^\circ$. Subjects used a chin-rest to stabilize their head. Data were acquired from the right eye alone. All subjects had normal or corrected-to-normal eyesight.

¹⁶ See <http://www.klab.caltech.edu/~moran/db/faces> for the entire list of the images used in the study.

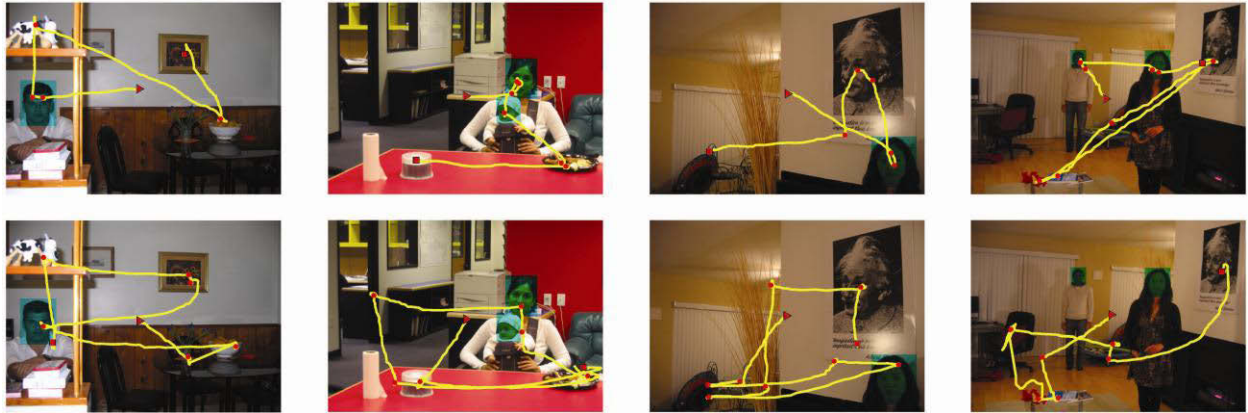


Figure 80 - Examples of stimuli used

Notice that faces have neutral expressions. Upper panels include scanpaths of 4 different controls, while low panels are of autism subjects. The red triangle marks the first and the red square the last fixation, the yellow line the scanpath, and the red circles the subsequent fixations. Region of interests used for the analysis are highlighted in the images. The trend of visiting the faces first — typically within the first or second fixation — is evident for the controls, while autism subjects visit the faces later and tend to explore other noticeable objects in the scene.

Experiments were approved by the Caltech Internal Review Board. All subjects signed an informed-consent form prior to the experiment.

Subjects were diagnosed as belonging to the autism/control groups based on a preliminary interview and a set of tests.

All autism participants were adults with high-functioning autism or Asperger's Syndrome, according to DSM-IV criteria determined by a clinical diagnosis. They had minimal co-morbidity (nobody had major depression or other psychiatric illness), no neurological disease

(no seizure disorder or other evidence of neurological illness), and all had participated in other research studies on autism. All control participants were psychiatrically and neurologically healthy individuals without a family history of autism.

	ID	SEX	Age	IQ	ADI-R			ADOS			Benton	STAI	
					Social (cutoff = 10)	Communication (cutoff = 8)	Stereotypy (cutoff = 3)	Social (cutoff = 4)	Communication (cutoff = 2)	Stereotypy		State	Trait
Autism	1	M	35	99	25	15	5	3	6	3	50	20	28
	2	M	45	126							43	45	59
	3	M	42	11				4	5	1	47	31	42
	4	M	36	128							49	46	59
	5	F	18		25	18	3					45	43
	6	F	18		24	18	4					36	29
	7	F	32	133							47	59	49
	8	M	24								47	34	39
	9	M	19	96	23	18	9	5	9	1	37	27	50
	10	M	26	106		21	20	6	17	0	49	52	34
mean			29.50										
±			±										
s.d			9.96										
Controls	1-15	11 Males	27.47 ± 7.62	115.71 ± 8.54									

Table 7 – Information about autism subjects

Results

To evaluate the results of the 25 subjects' viewing of the images, we manually defined minimally sized rectangular regions-of-interest (ROIs) around each face in the entire image collection (See example of ROI in Figure 80). We first assessed, in the "free-viewing" phase, how many of the first fixations went to face, then how many of the second, third fixations and so forth.

Controls attended to the face ROI on average in $67.56\% \pm 8.93$ of the trials within their first fixation, while autism subjects attended the face only $54.4 \pm 4.11\%$. While controls visited the faces above 90% of the times by their second fixation, autism subjects reached the same level only by their fourth fixation. This suggests a delay in the attention to faces across the two groups. Testing for the similarity in distributions indeed shows that they are significantly different ($p = 0.02$, Wilcoxon). Figure 81 shows the distribution of first fixation across the two groups in the "free-viewing" task.

In order to disentangle the ability to look at faces early, from the voluntary will to attend to those we had subjects participate in a second experiment where they were encouraged to look at faces first in one half of the trial (search for a face), and encouraged to look elsewhere in the other half. If autistics are able to look at faces first, and are electing not to, then we expected to see higher similarity in the first fixations distribution for this experiment. Indeed, when comparing the results of the first visit to the face ROI in the second experiment for the two groups the difference between the two distributions is no longer significant. While controls visit the faces $16.3\% \pm 7.01$ in their first fixation and reached only 29.1% by their second fixation,

autism subjects visit faces in $18.9\% \pm 4.91$ and reach 31.57% within their second fixation. Both groups show a drop in the amount of fixation allocated to faces altogether. See Figure 81 for an illustration of the distribution for the search task. This suggests that indeed the two groups are indistinguishable in their fixation patterns, given a clear task directing their attention. Figure 81 shows the individual proportions of first fixations for the two groups and the two experiments.

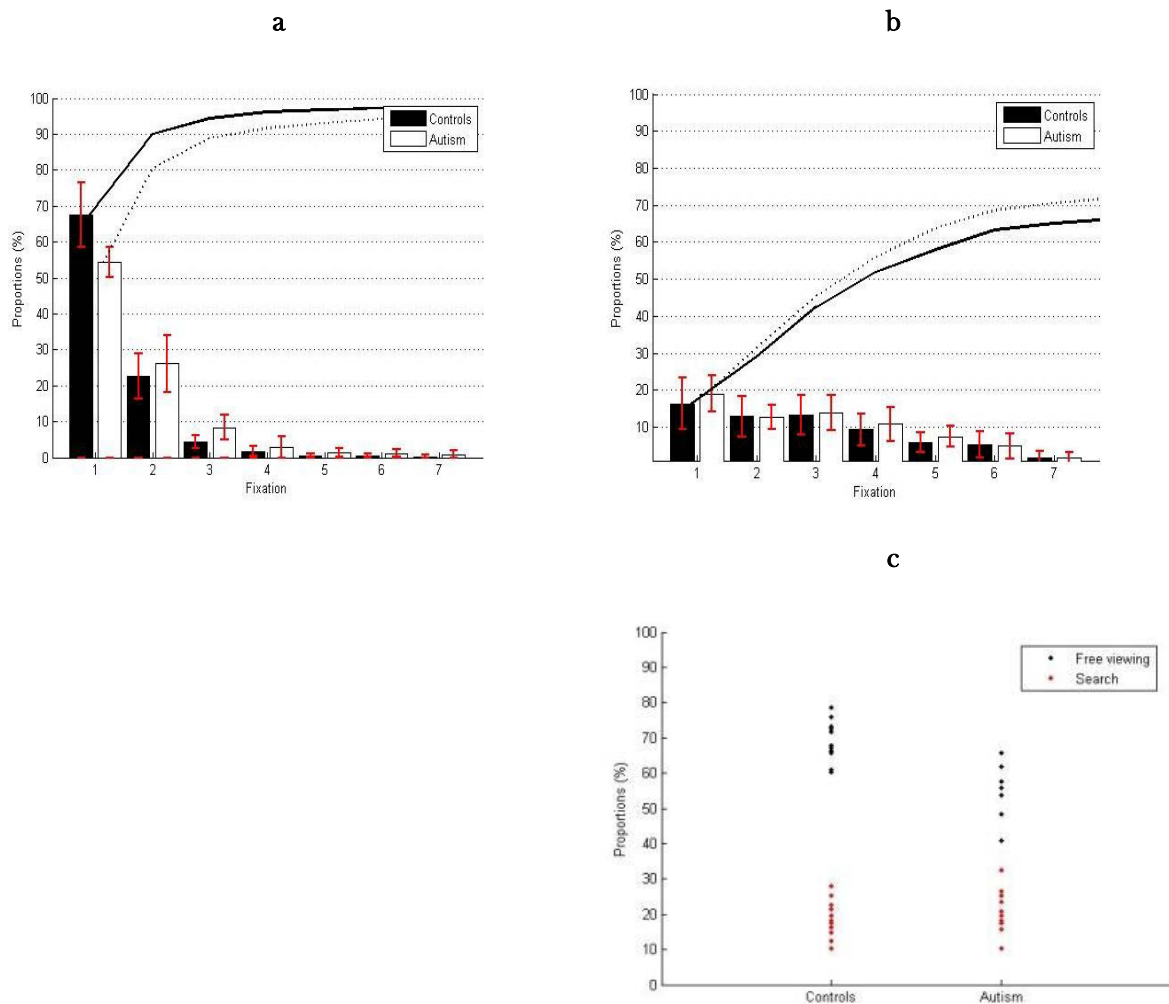


Figure 81 – Proportion of first fixations in autism

a. Proportions of first fixation on the face region of interest for the autism and controls groups, in the free viewing task. Results are normalized. Red bars indicate error of averages over 10 (autism) and 15 (controls) subjects. Bars depict percentage of trials, which reach a face the first time in the first, second, third... fixation. The solid curve depicts the integral, i.e., the fraction of trials in which faces were fixated on at least once up to and including the n th fixation.

b. Proportion of first fixations on the face region of interest for the autism and control group, in the search task.

c. Individual subjects' results for the first fixation in the two experiments. The left column of dots shows a significant separation between the upper cluster (free viewing) and the lower cluster (search). Right column shows the proportion of first fixations for the autism group.

Quantifying the time it takes either group to attend to the face suggests that while autistics do sometimes attend to the face in their first fixation, this takes a longer time than it takes normals (Controls: $297.82 \text{ ms} \pm 48.85$, Autism: $355.90 \text{ ms} \pm 77.92$). This suggests that while autistic subjects do look at the face sometimes, they do that more consciously (based on more of a learned skill) rather than being automatically saccade driven by an internal bottom-up saliency mechanism. Interestingly, the time differences are comparable between the two groups in the search task when they are asked to find the faces, and is much longer for either group (Controls: $760.61 \text{ ms} \pm 131.47$; Autism: $758.70 \text{ ms} \pm 159.72$). The latency differences for the free viewing task are significantly different ($p = 0.01$, Wilcoxon), while they are not significant for the search task.

In order to better understand the underlying tendency to look at faces regardless of the task, we looked separately at two categories within the search task: a) trials where the face was visited when subjects were instructed to find an object in the image, and b) trials when the face was visited when subjects were instructed to find a face. In trials where the target was a face (subjects are therefore encouraged to look at a face faster) controls attended to the faces after $707.80 \text{ ms} \pm 462.02$ while autistics took $740.59 \text{ ms} \pm 511.61$. In contrast, in trials where the target was an object, controls looked at faces within $817.33 \text{ ms} \pm 443.79$ while autistics spent $780.50 \text{ ms} \pm 501.85$.

Testing for the inter-subject consistency within group and across groups in the viewing patterns for the given images shows that the variability across the two groups is the same upon looking at the images in either experiment (Kruskal-Wallis). However, χ^2 test for the proportion of first fixation to faces versus the remaining fixations for either groups, shows a significant difference between the two groups in the free viewing experiment ($p = 0.03$) and in the search task ($p < 10^{-5}$, Fisher exact test).

Although the images were the same for both groups, which suggests that the difference in fixation patterns is indeed due to attention selection, rather than attributes of the images or the size of the face relative to the image, we verified the results by normalizing the fixation patterns for the relative size of the face compared to the entire image. This was done by creating an unbiased baseline estimate for the amount of fixations in a face at random. The baseline for a particular image is the fraction of all subjects' fixations from all other images that fall in the ROI of the particular image. This baseline also controls for the center-bias, and the

photographers' tendency to locate relevant social cues at the center of the image. The null hypothesis that we would see the same fraction of first fixations on a face at random is rejected at $p < 10^{-10}$ (t-test) for the normals, and at $p < 10^{-8}$ for the autism group (Cerf, Harel, Huth et al., 2008).

While control subjects look at faces earlier than autistic subjects, the question of how interesting these are for them remained elusive. To test that we measured the amount of time spent in facial regions within the 2 s allocated for the task. Out of the 2 seconds, controls spent $1092.83 \text{ ms} \pm 183.86$ (54% of their time) attending to faces, while autistic subjects spent $1037 \text{ ms} \pm 352.12$ (51%) of their time in the free viewing task, and $547.40 \text{ ms} \pm 103.23$ and $598.65 \text{ ms} \pm 55.55$ (controls and autistics, respectively) in the search task. Results are not significant, suggesting that while the drive to attend to the face is different across the two groups, when these are visited they tend to elicit similar level of interest, measured by the amount of time allocated to exploring the faces.

As a measure of the subjective level of interest of each group in the content we also used the self-reported interest ranking subjects gave for each image in the dataset. While subjects were not instructed to distribute their answers in any way (they may as well find all the images boring, ranking them a 1, or finding all of them very interesting, ranking them a 9) we observe a trend of typical random distributions in subjects' answers, with tendency to find the later images more boring, which is likely to be due to adaptation to the category with time (Cerf et al., 2007). The mean ranking for the control group was 3.81 ± 1.80 , and the mean ranking for the autism group was 4.36 ± 1.57 (not significant differences, $p = 0.28$, 2-tailed heteroscedastic t-test). Separating the distributions to images with faces (controls: 3.89 ± 1.68 ; autistics: $4.29 \pm$

1.28) or images with no faces (controls: 3.58 ± 2.02 ; autistics 4.57 ± 1.55) shows significant results for the no-faces images ($p = 0.05$, 2-tailed t-test), but otherwise is not significant.

In order to test the degree of correlation between autism disorder and the patterns of attention allocation to social scenes we correlated subjects' viewing patterns with independent data collected with these autism subjects. We devised a measure of performance in the task by the proportion of faces visited by the second fixation (control subjects reach a level of 90.07% of attending to faces by that fixation) and for each individual subject correlated the performance with his ranking in independent tests (Benton, STAI, and IQ). The 3 tests' results had very little variability (and matched these of the controls), thus the correlation is insignificant.

In order to evaluate the performance on an objective task to make sure that the difference between the two groups are uniquely due to the social content of the images and not other impairment we used two independent experiments. First, we tested the performance on the memory task. Subjects were asked to correctly recognize 100 images that they previously saw in the experiment. This measure allowed us to test the attention, making sure that indeed the subject noticed the images seen, as well as test the level of attendance by subjects. Indeed, both groups reached virtually the same perfect results in the task (controls: 97.47 ± 2.42 ; autism group: 97.40 ± 2.41). This suggests that subjects were very attentive during the experiment and were looking carefully at the images at all times.

As a secondary measure we used another independent experiment. While faces in this study were assumed to be a social cue that distinguishes between autistics and controls, we used a neutral cue that is important and relevant for both groups, yet does not necessarily contain social relevance: text. Humans are exposed to text a great deal and very often in our lives, yet it

is very unlikely that text serves as a cue that tells autistics from controls. Prior studies of attention to text versus faces show that, while text is important, its role in attracting attention is significantly smaller than that of faces. We tested the controls and autistic groups in viewing a sequence of images containing text and no other significant social cue (no faces, people), in a “free-viewing” task. We evaluated the performance of the two groups similarly to the way by which we evaluated the faces’ ROI. The results for the text free-viewing in controls for the first fixation on a text object shows that controls visit the text ROI 22.9% of the time in their first fixation, and 19.1% in their second fixation. By their third fixation controls will have visited 71.8% of the text regions of interest. Results for autism subjects show similar patterns. Autism subjects visit the text 21.1% of the time within their first fixation, 17.1% within their second fixation, and will have visited 70% of the text by their third fixation. There is no significance difference between controls and autism for free viewing of text images. Interestingly, the results of the autism group viewing face images are very similar to those of controls viewing text. This supports our hypothesis that while faces are very economically important for the autistic subjects, they are not drawn to them naturally, but are rather trained or habituated to attending to them based on practice and experience. While controls are suggested to have a bottom-up driven saliency mechanism alerting their attention and drawing it to faces covertly, the autism group is attending to faces based on experience and training (hence, the longer latency in attending to those in the free viewing task), and as such show similar patterns of viewing to those of controls attending to text which is an important cue, that is not social, but is trained to be important. Similar measure of the results using an unbiased baseline (all fixation of an individual subject for all images used as a measure for his attention to a particular location) show that the similarity between the viewing patterns is even more pronounced given the

individual biases. See Figure 82 for the distribution of first fixations across the two groups and the new text region.

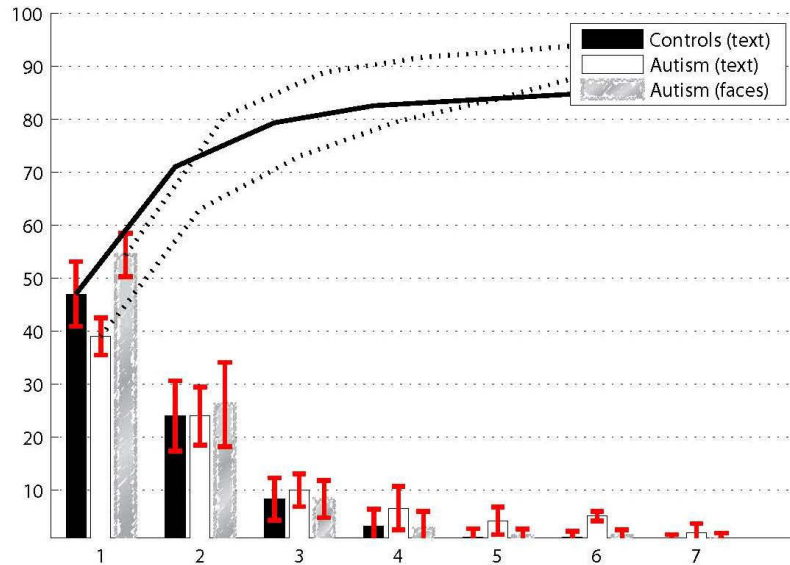


Figure 82 – Proportions of first fixation in the region of interest

Solid bars show the results for the text ROI, while the scribbled bars show the results for the faces (“free viewing” task, as shown in Figure 81). Red lines indicate error bars. Bars depict percentage of trials, which reach a face/text the first time in the first, second, third, etc., fixation. The solid curve depicts the integral, i.e., the fraction of trials in which faces/text were fixated on at least once, up to and including the n^{th} fixation. The distributions of the autism group viewing faces resemble those of the controls viewing text.

Discussion

We quantified the amount of attention allocated to social cues (in this study faces) as a measure of level of interest and relevance social entities form for two groups of individuals – one with autism and one control. The autism group is highly functioning (see Table 7), yet are showing significant differences in their viewing patterns of faces given a free-viewing task. One can suggest that a higher level of autism may result in even decreased attention to faces.

We found that while in both experimental conditions (“free-viewing” and “search”), faces were powerful attractors of attention in normals, accounting for a strong majority of their early fixations when present, the trend is not similar for individuals with degrees of autism. This group shows significant difference in its attention allocated to faces, showing decreased bottom-up driven fixations towards those. Autism subjects look at the faces later and slower if at all, as if their overall mechanism gearing the attention towards the face is less developed and dependent on the instructions and task. While normal subjects look at the faces early independent of the task, almost unable to avoid looking at these even if that is not economically useful in order to succeed in the task, the autism subjects in fact change their fixations allocation pattern based on the task, and improve their performance if specifically asked to look at the faces. This suggest that faces are not as important for the autism subjects, and supports the hypothesis that a major difference between autistic and normal subjects lies in an undeveloped social saliency mechanism that shifts the attention rapidly to the most socially relevant entity in an image (faces, in our case).

The comparable results in the search task show that indeed both groups are able to perform the task in a similar way if need be, but when given the freedom (“free viewing”) to allocate their fixations based on their interest show significant difference.

Breaking the search task timing analysis to a) look at faces when instructed to look at objects versus b) looking at faces when instructed to look at faces; shows that the control group is 110 ms faster in looking at face when told to look at those, while the autism group is only 40 ms faster. This suggests that the autism group regards the two types of instructions much more similarly than to the controls. Finding an object or finding a face is a similar search task for autism, while for controls it is a very different one, yielding much higher differences in timing and proportions of first fixations to faces.

Further analysis of the two distributions in fact shows a similarity between the way autism subjects look at faces and the way by which normal subjects look at objects, suggesting that the social value of the object is of no import for the autism group (faces are just yet another object in the image), while they are very socially relevant for normals. This suggests that autism subjects are looking at the face in a similar manner to which normals look at an object: they can look at it quickly and even in their first fixation, but it is not comparable to the bottom-up driven way – rather it is a top-down task-driven one.

This eye-tracking experiment sheds light on a difference in the social value autistic and normal subjects attribute to faces – highly socially relevant cues – in a scene. This is also reflected in the correlation between the performance in looking at faces and the BOLD signal for the two groups.

These clear differences in looking at social cues between autistic and normal groups can be used in a controlled manner to even add an additional diagnostic measure to quantify the degree of autism along the spectrum (Cerf, Harel, Huth et al., 2008).

Finally, similarly to the way by which normal subjects' fixation allocation could be modeled using a simple saliency algorithm combined with face detection, we can identify the "entity" which results in similar performance for the autism group and devise a saliency algorithm for autism subjects that might shed some light on the ways by which these individuals judge a social scene. More so, using the recently introduced "decoding metric" value, we can correctly identify a subject coming from the autism or control group in a classifying manner.



Anecdotal cases

A collection of anecdotes is the greatest of treasures for the man of the world, for he knows how to intersperse conversation with the former in fit places, and to recollect the latter on proper occasions

Johann Wolfgang von Goethe

Monozygotic (identical) autistic twins

Introduction

Monozygotic twins share many behavioral traits. The chances of such siblings both being autistic are higher than the average. In the course of my work I encountered a single case of two identical autistic 18 years old twin sisters, upon whom I had the opportunity to run a battery of attention tests. Following is a summary of an experiment done with the twins. It all started when I was told that the two twins are not particularly talkative and that I should pay careful attention to their behavior during the experiment, as they won't necessarily tell you that they are unhappy when they are. I decided to take frequent breaks between tasks and tried to get them to interact with me in order to make sure they were engaged in the experiment and weren't unhappy about it.

EM was the first sister to run through the experiment. In order to make conversation I had her take her first break at the end of the first block (after seeing 200 images out of the 760), and asked her general questions about the experiment: "what type of images did you like more?", "Which ones did you not like?", or general questions about herself: "What do you like to do for fun?", "what classes do you like in the university?", etc. She gave me an assortment of answers that I didn't think I should pay careful attention to at this point as they were merely asked in order to keep her alert and happy.

About halfway through the experiment (fourth block, 400 images seen by that time) SM (sister, not to be confused with the amygdala-lesion patient SM) and their mom joined us in order for the 3 to go have lunch. The mother showed a great interest in the experiment, and so I chatted a bit with both the mother and the SM while EM was finishing her fourth block. I asked SM a few questions about her interests and school and it was very clear that the answers were exactly the same. I pointed out to the mother and to SM that the answers are surprisingly very similar to those of her sister, and the mom said that it was not surprising to her, as they show these similarity issues regularly. I pointed out that it was interesting and that it would be even more interesting to look at their responses to the images seen. That's how things started...

Method

While typically we don't look at the "answers" subjects give to the questions asked during the course of the experiment (mostly answers to the question: "How interesting was this image (1-9)?"), as we care more about the scanpath and regard those as merely a way to keep the subjects alert during the course of the experiment and to control for the task subjects performance being independent of the content, we do keep a set of 200 answers to the same question for each of blocks 1, 3, and 5 and another 80 images for block 7 and 8, giving $200 \times 3 + 80 \times 2 = 760$. (In blocks 2, 4, and 6 we ask different questions – either questions that have to do with the emotional judgment of the images (valence/arousal), or questions that have to do with search task in block 2: "Did you find the object in the image?")

So for the immediate analysis I just collected the set of 760 numbers that came from a shuffled sequence of images presented to either sister and quantified the agreement between the two. I used two metrics for that – at first, I just ran a simple correlation between the two streams of 760 numbers, and then I also quantified the amount of **exact** same answers to the images seen.

Results

Correlation

The correlation between the two sisters at first was 0.71. Obviously I was intrigued and had to quantify the agreement between two random pairs as chance. It turns out that on average the agreement between two random pair of subjects is: 0.39 (this is in agreement with our earlier Vision Research study where we show that the agreement (correlation coefficient) between subjects viewing images of natural scenes is about 0.3 (in the previous study the colored natural scenes were a little less intriguing as they did not contain faces and known people, which can account for the minor 0.09 difference in number).

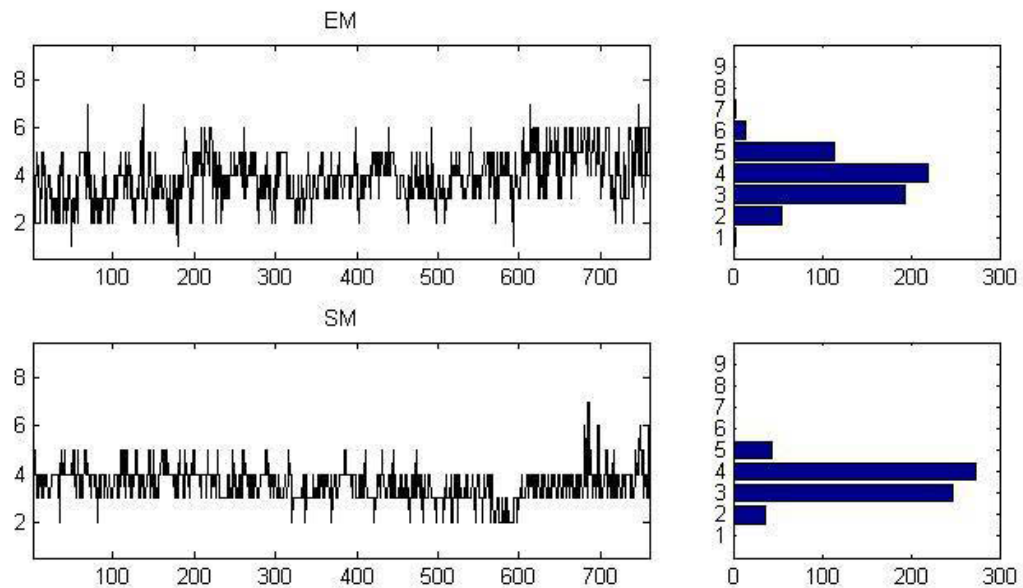


Figure 83 - Rating of images for the identical autistic twins

On the left panel, the sequence of the 760 images with their rating for both sisters. On the right, a histogram of the answer distribution showing that both sisters mainly used the middle of the scale and didn't find any image extremely exciting. (EM: 3.93 ± 1.09 , mean \pm s.t.d.; SM: 3.59 ± 0.75)

Categories

Now, while those numbers pertain to the entire collection of 760 images, if I broke the sets into to blocks (which makes sense, as the 760 images are coming from different categories: faces, text, anomalies, IAPS, black and white, etc) I could see that for the typical color faces images the correlation was even higher (0.91 for a set of 350 color images with faces from all sets). This of course is significant ($p < 10^{-20}$) where the null hypotheses is that the streams come from random numbers, and is significant if you compare it the average agreement between randomly chosen pairs of subjects (that is, by itself above chance already). Remember that those images were seen in random order by each sister and as such are in fact telling us that it's the image rating that is correlated and not the order or the tiredness or any factor outside of the answers themselves. So, this was good news.

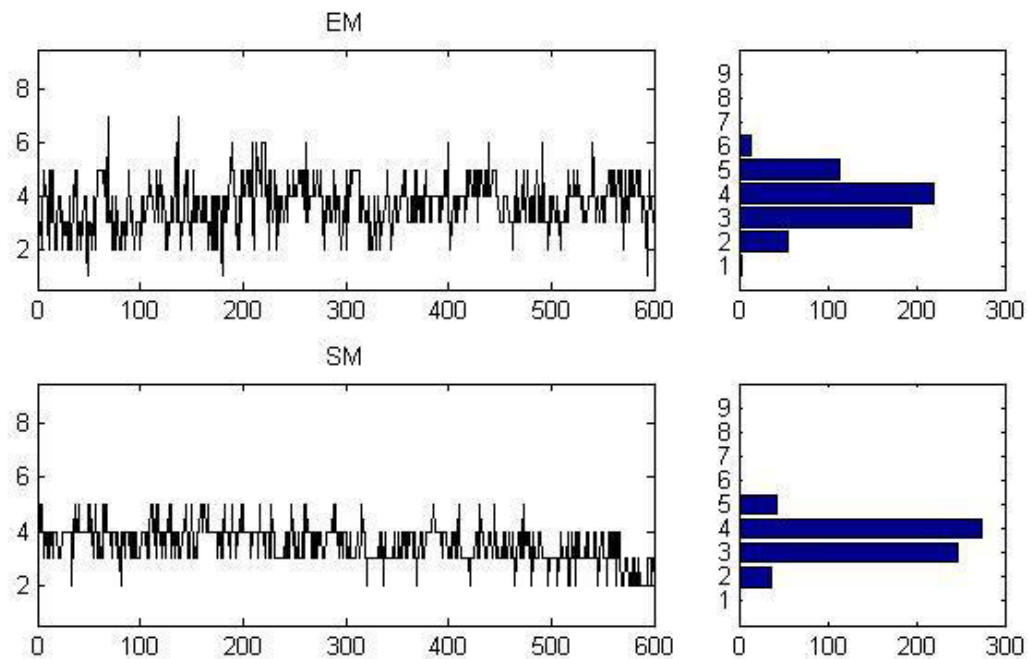


Figure 84 - Rating of the entire dataset for the identical autism sisters

Ratings for 600 images that include natural scenes

Correlation without sequences

However, this in fact was too good to be true — and early on I figured out why. It turns out that for many images out of the 760 seen by both sisters in fact used a very limited amount of numbers out of the 1-9 image scale offered to them. They in fact used mainly the numbers 4 and 5 for their judgments of the images. This yields a bias towards an agreement between the two that leads to the higher correlation and we see. (Although one can argue that using a limited number in order to judge the images seen is still a valid answer to the open question “how interesting is the image?” That is, if both sisters find most images equally boring or interesting then that is a valid result, and we can’t really claim that it is necessarily a bias in our results.

In order to control for that bias I decided to use two methods. The first takes out all of the images where the sequence of answers given by either sister followed a set of 5 or more repetitive numbers.. That is, if she gave answers of 5-5-5-5-5-5-5-5-5, then only the first 4 (5-5-5-5) would be taken and I would discard the remaining answers. This reduces dramatically the amount of answers we get, but still yields a statistically significant result. The second quantifies the effect of the large repetitive sequences on the results by creating 1000 random shuffles of the same numbers (so the histograms remain the same, but the order is broken), and measuring the correlation between those compared to the correlation of our shuffle. If they are not different then it means that the repetitions are not the source of the higher correlation. Shuffling the sequence and running the correlation shows that the average correlation for 1000 numbers is 0, with 3 random correlations falling above the real one — suggesting a p value for the expected correlation at random of 0.003.

Perfect agreement

If one just counts how many images they agree upon exactly — meaning: both sisters gave EXACTLY the same number out of 9, and you discard sequences of more than 5 repetitive answers in a row you get 30% agreement. Chance here is given by a binomial function with $p = 1/9$ for match, $n = 600$ trials and $k = 182$ (perfect matches). I needed to work out a numerical approximation for this function, as factorials of that magnitude are obviously too big for Matlab

to process, but if you use the log of Gamma function (which estimates factorials), then it would be approximately (lower boundary): $p = 2.5 \times 10^{37}$.

This, however, assumes that the distribution of answers is in fact unified (meaning they potentially can use the entire set of 9 number). If we want to be more conservative and really tie the results to the twins then we can maybe narrow the scale to fewer numbers, as the probably of choosing 9 is lower than the probability of choosing a 4 in their case.

Clustering

Trying to cluster the images sequences by the ones for which they had the largest correlation (black and white, with faces, with objects, text, etc.), it seems that of all the classes of images viewed by the two, the category that elicited the highest correlation was the one where they viewed “anomalous” images (images where we inserted a deliberate anomaly like removing the eyes from the face, rotating parts of the face, etc.). Correlation coefficient for those is $r = 0.35$ ($p < 10^3$).

Correlation for the IAPS images alone is $r = 0.29$ ($p = 0.05$).

IAPS

For the IAPS images I also checked the actual ratings they gave for valence and arousal, and it seems that the numbers here are also correlated. For the valence it's $r = 0.48$ ($p < 10^{-7}$), and for the arousal it's $r = -0.25$ ($p = 0.01$). That is, for arousal they are in fact negatively correlated.

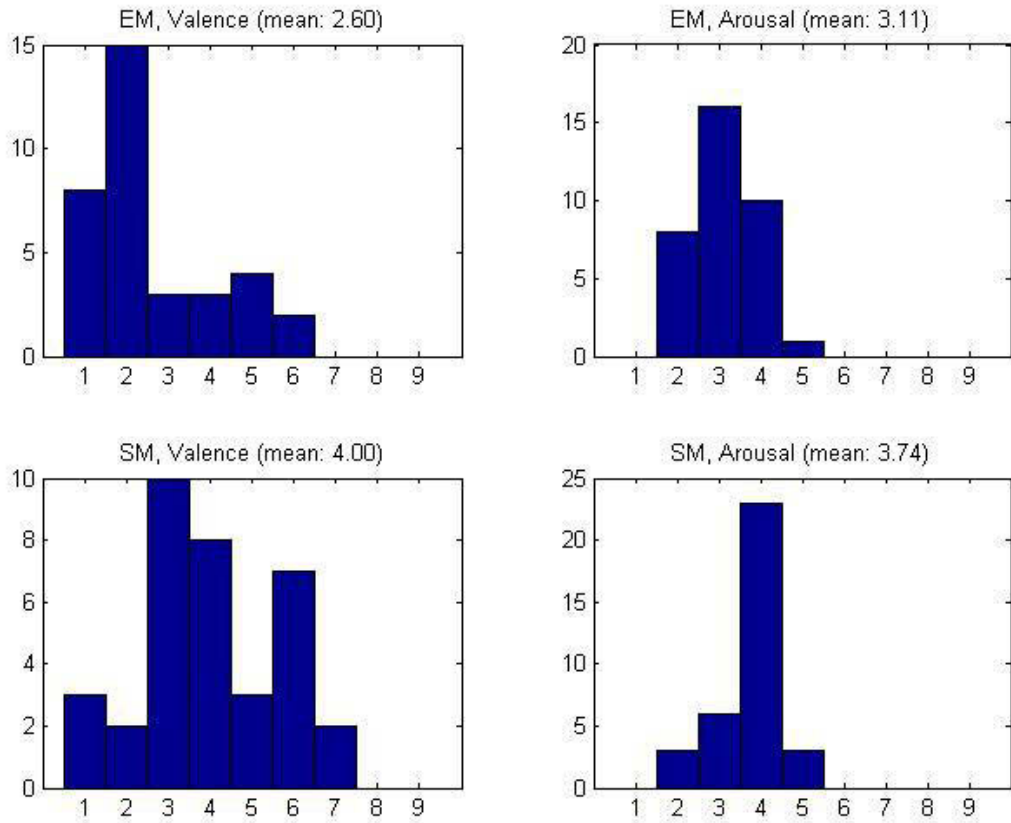


Figure 85 - Distributions of answers

The IAPS block (100 images) distribution of answers for both sisters

Eye-Tracking

As for the eye-tracking data, while it is harder to quantify the agreement between 2 scanpaths of data, plotting the raw data suggests that EM was having more fixations and was scanning the environment more than SM, although judging by their early fixations the scanpaths look as if the twins had indeed continuously looked at the same objects throughout their scans.

Interestingly, we noticed that it seems that their eye-tracking data shows very similar scanpaths, even for images without faces. This is not typically the case for other pairs of subjects.

See Figure 86 for examples.

Take a look at images 055, 059, 072, 078, 080, 081, 089, and many other images with no faces.



Figure 86 - Examples of scanpath for the autism identical twins

Right. Scanpath while viewing image 34 in block 1 (“free viewing”) for EM. Image rating was 2.

Left. Scanpath while viewing image 34 in block 1 (“free viewing”) for SM. Image rating was 2.



Figure 87 - Examples of scanpath for the autism identical twins

Right. Scanpath while viewing image 54 in block 1 (“free viewing”) for EM. Image rating was 4.

Left. Scanpath while viewing image 34 in block 1 (“free viewing”) for SM. Image rating was 3.

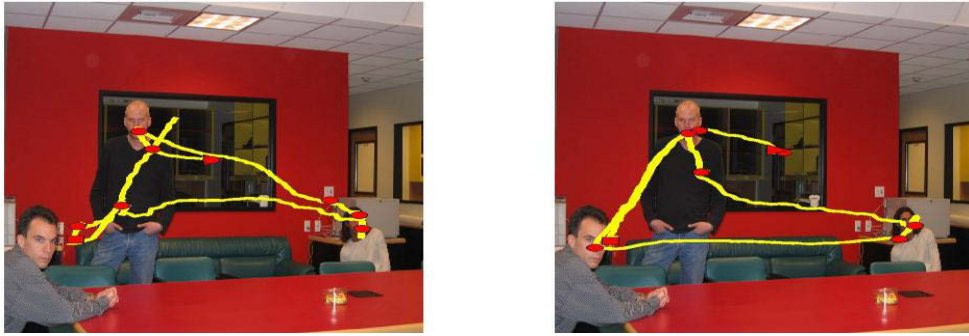


Figure 88 - Examples of scanpath for the autism identical twins

Right. Scanpath while viewing image 186 in block 1 (“free viewing”) for EM. Image rating was 4.

Left. Scanpath while viewing image 186 in block 1 (“free viewing”) for SM. Image rating was 4.

Conclusions

While the results are very interesting and might shed light on a bigger phenotypical correlation that highlights a genetic relation, these are very rare subjects and are hard to test. This falls under a specific study that one can do on the perception of images by subjects, suggesting a higher correlation between people with shared background, which most often is the case among twins. While we already demonstrated in chapter 9 that a “general walk” in the museum where one supposedly ranks the images independently” is in fact more similar to others’ responses than we think, now we can extend that claim and say that in fact if we are speaking of two identical twins, we can just send one of them to attend the museum – as the other would just judge everything the same anyhow.

Prosopagnosia

I had the privilege of testing two unique subjects with high level of prosopagnosia. Subject **KM** is a 25-year-old graduate student with prosopagnosia, while subject **MH** is in his 40s and has a higher level of Prosopagnosia accompanied by various other vision disorders (inability to tell objects from background in far scenery, and very limited peripheral vision). **KM** has no known brain conditions leading to the face blindness while **MH** suffered a brain trauma at age 5 which led to the various vision disorders after being in a coma for months. Notice that while subject **MH** is “well-tested” and was studied numerous times for various brain, vision, and psychology disorders, subject **KM** never went through a brain scan, and as such need to be tested further in order to see if it was possible to identify or isolate a neurological disorder leading to the face blindness.

Both subjects participated in the standard eye-tracking experiments where they were viewed images with faces and social scenes and were asked to either report how interesting the images were on a scale of 1 to 9 (“free viewing”) or locate objects/faces in the images based on a previously viewed probe (“search task”).

Results

Testing for the proportion of first fixations in face regions in the free viewing task for both subjects show conflicting results. While subject **KM** shows results identical to these of controls (Figure 89) subject **MH** shows a dramatic reduction in the ability to look at faces. Subject **KM** missed only 2 faces in the course of the entire experiment, and generally viewed nearly 70% of the faces in the first fixation, while subject **MH** missed 30 faces and his results fall short even for the lowest results for autism or AgCC subjects, suggesting indeed a failure to identify the mere existence of a face.

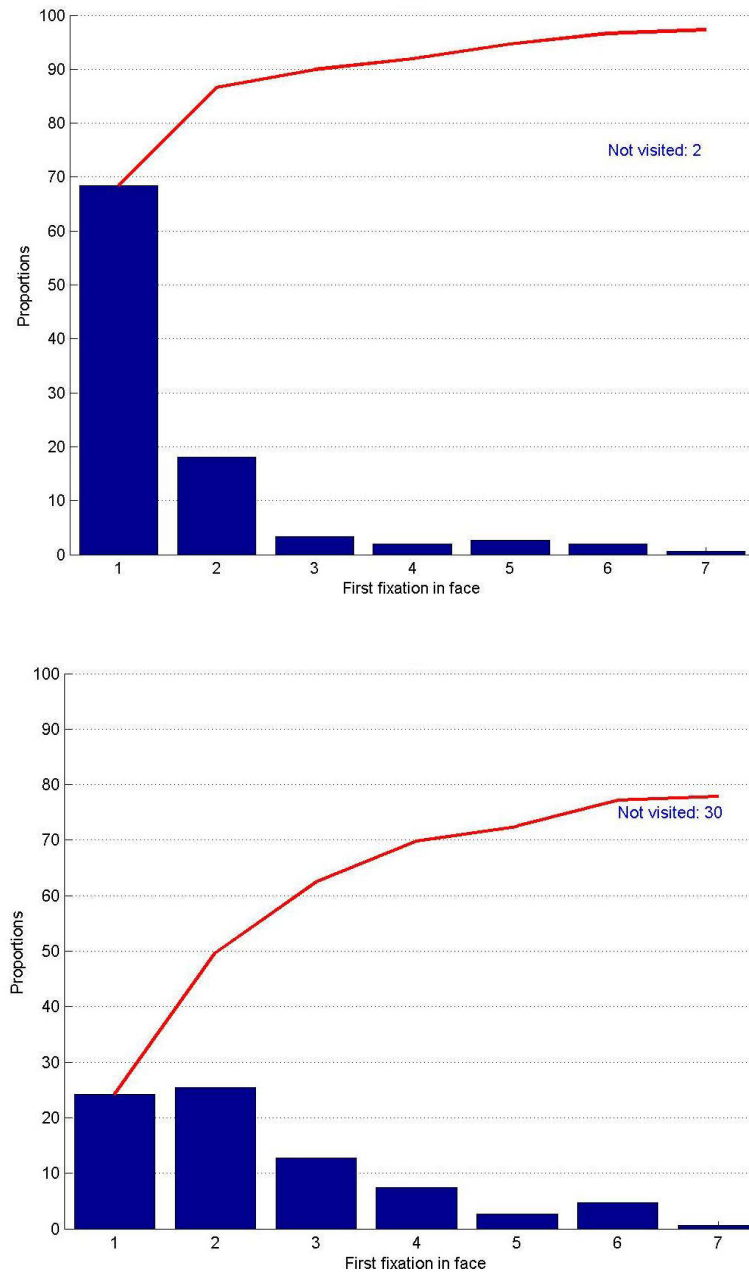


Figure 89 – Proportion of first fixations on a face in “free viewing” in prosopagnosia

Top panel. Subject KM, who shows mild prosopagnosia, looks at faces fairly early (visiting 90% of the faces by her third fixation).

Bottom panel. Subject MH, who suffers from severe prosopagnosia, shows a dramatic decrease in his ability to visit faces at all (30 faces – which are about $\frac{1}{5}$ of the faces in the entire set – were not visited at all. He, therefore, never reaches 90% throughout the entire experiment, and visits only 25% of the faces in his early fixation.

Testing the two subjects' results in the search task in terms of their ability to look at faces when searching shows no significant differences from the results of controls performing the same task (Figure 90).

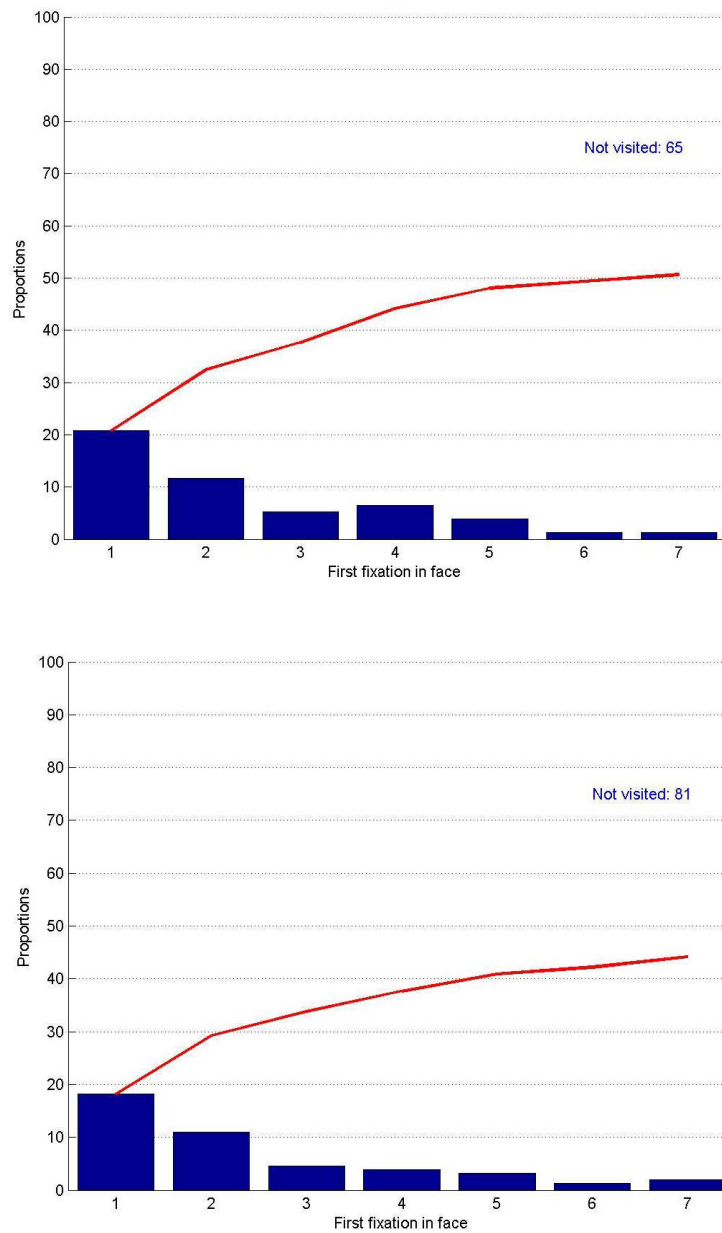


Figure 90 – Proportion of first fixations on a face in “search task” in prosopagnosia

Top panel. Subject KM proportion of first fixations on the face for the search task. Her results are within the range of controls.

Bottom panel. Subject MH proportion of first fixations on the face for the search task. His results are within the range of controls.

While the two subjects show no difference in their ability to locate the face in the scene, which is the exact task at hand, they interestingly show a very significant reduction in their performance in the task with respect to accuracy. For previous subjects performing the task we never tested the amount of correct viewing of the face as the task is made such that the probe is first shown and then the assumption is that actually identifying it when fixating on it is a given. However, for the prosopagnosia subjects this was the majority of the hardship. (See Figure 91 for an illustration of the task and its problem for subjects with prosopagnosia.) Out of all the images that had multiple faces in them in the search task, subjects correctly looked at the target face 77% of the time, while controls look at the correct face 100% of the time ($p < 10^{-20}$).

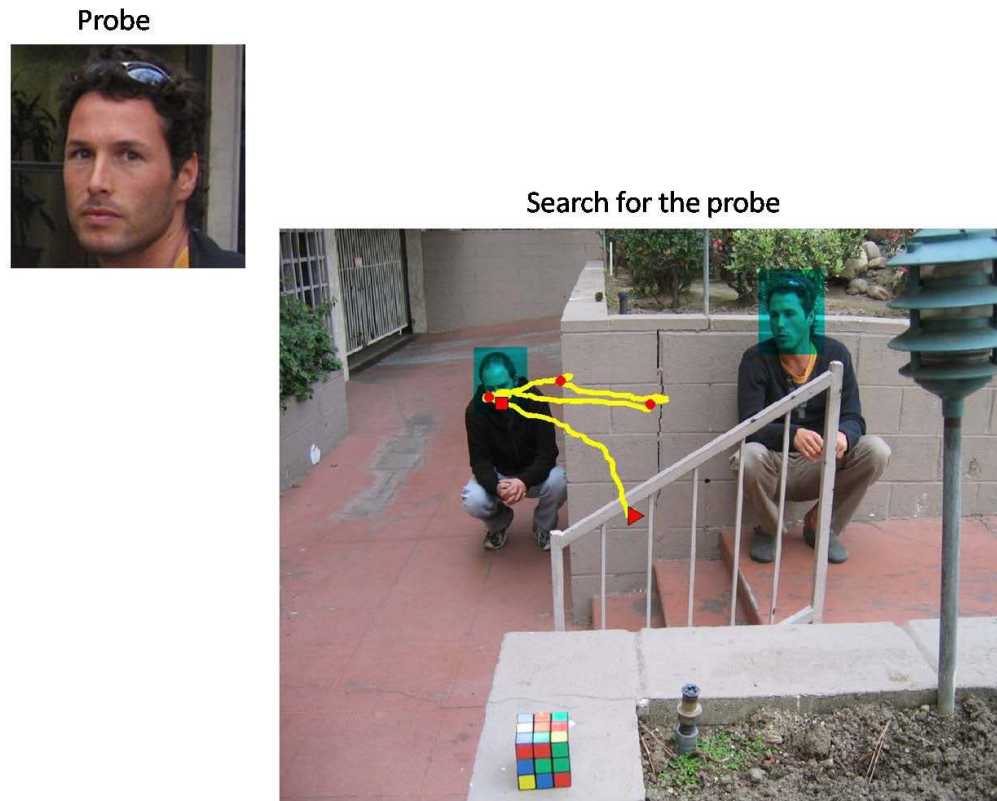


Figure 91 - Failure to correctly identify the face in Prosopagnosia

In the particular example the target (shown for 600 ms prior to image presentation) was of a person named Yadin (shown in the image on the right). The prosopagnosia subject looked at the face rapidly and within his first fixation, as shown in the results for controls in previous chapters. However, the subject in fact looked at the wrong face, and completely ignored the correct one. This result would not be reflected in our analysis, which looks only at the first fixation on any face. That said, most of our images contain only a single face and therefore would not be susceptible to this mistake.

As we did not design the experiment for this particular case, we could not alter the dataset in particular for these subjects in order to measure this effect in particular, and the results are

purely anecdotal – yet they go hand in hand with our belief of how prosopagnosia subjects function – they can see a face, but they cannot identify it. These results suggest that while they do see a face, they don't share the weighting systems towards its being an attractive prospect in the environment, making them attend to faces less even when they are aware of their existence.

Discussion

Subject MH's not looking at faces could arise from one of two main reasons: he could either have not identified a face at all, and as such not visit the faces, or he could have identified the face but not share the attention mechanisms that drive most of us to look at a face. In the competition context we can imagine the information about the face flowing in his brain, but suppressed by other objects in the environment that repeatedly win over his attraction. Since subject MH is able to see faces, and his results from the general prosopagnosia diagnostics exams show that he is capable of looking for and identifying the existence of a face, I tend to support the second hypothesis regarding his lower attraction to faces. This could imply that areas in his brain that should be influential in possessing the majority of the attention are in fact not doing so – allowing for various other mechanisms to pull the saccade towards objects in the environment that are not faces. Testing for the attention to other objects from the pool of objects we had manually defined regions of interest for as the banana, or the Rubik's cube show no significant heightened attention. That is, it doesn't look like MH is looking particularly at other things. For all I can say, he scans the environment as if the faces are just another artifact in the scene – based mainly on his own interest combined with the saliency of objects. Indeed, correlating his results with a saliency model with no face detection (based purely on features in the images such as color, intensity, and orientation) compared to saliency model with face detection shows that the first does better in predicting his fixations. This is the only subject for which this is true. Even the lower functioning autism subjects or AgCC subjects show an increased performance for the saliency model with faces compared to the one without. Interestingly, MH's performance looks very comparable to the results of controls looking at text and objects (that is the text still reaches a higher fixation proportions of about 50% in the

first and second fixations combined, supporting the suggestion that MH indeed has the ability to identify faces but they just do not capture his attention when competing for it with other cues from the environment.

Amygdala lesion

SM was visiting Caltech for one week. During that week I spent enormous amount of time studying her through various studies. I will report here the results from each of these studies.

Experiment I – Attention allocation to faces

Conducting the standard eye-tracking experiment where SM is asked to look at images with faces and social scenes, rate their interest, or search for probed targets – replicating the results from chapter 10 did not show any significant differences in the results between SM and these of healthy controls.

SM was looking at the faces as much as the controls were. See Figure 92 for histogram of her first fixation in faces. Overall, SM visited all the faces shown to her in the images and reached a threshold of visiting 90% of the faces by her second fixation (as did controls).

In the second – search task – SM again shows a behavior similar to that of the control subjects, reducing her level of viewing of faces dramatically (see Figure 92).

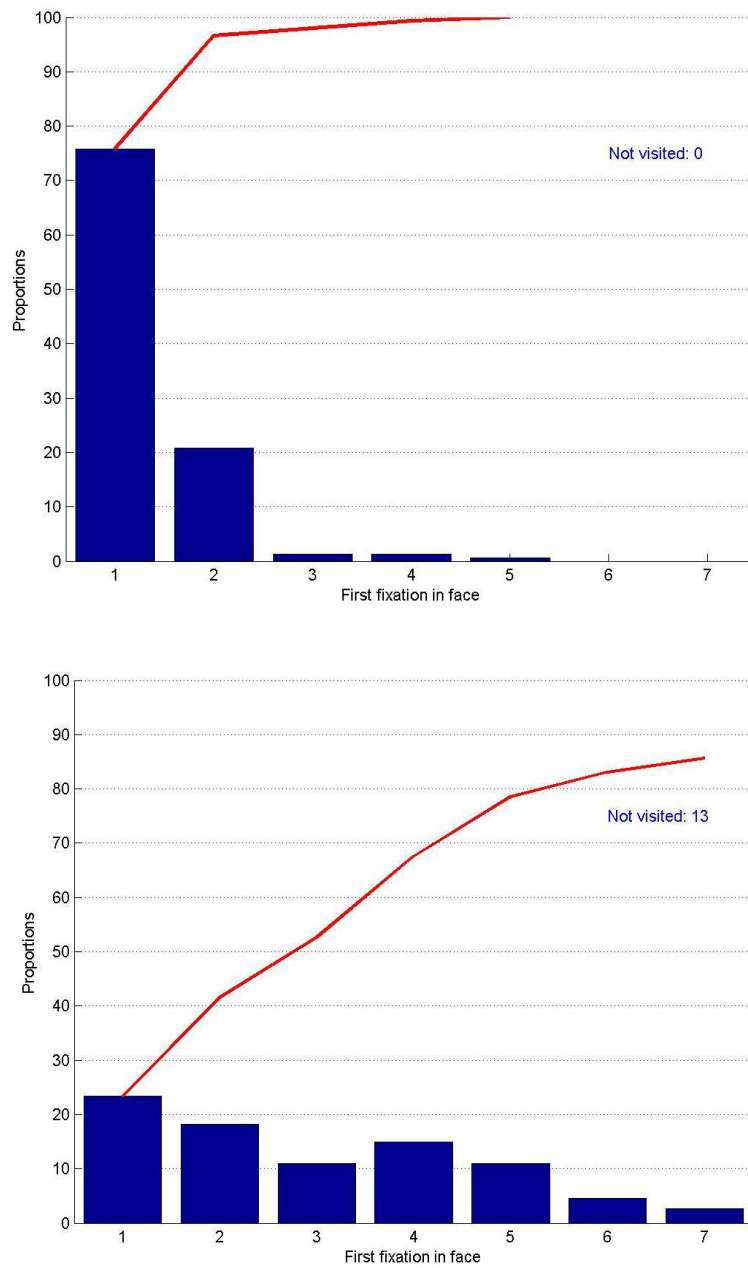


Figure 92 - SM looking at faces in blocks I and II

Top panel. SM looking at faces in the first (“free viewing”) block. The results fall within the range of healthy controls and SM does not seem to show any significant difference in the proportion of first fixations on the face.

Bottom panel. SM looking at faces in the second (“search task”) block

While there is a room for more thorough comparison of SM’s results to those of controls (in terms of timing and further metrics such as number of fixations, time spent on face regions, etc.), none of these were conducted at this stage, and the nature of the results appearing similar to these of normal controls made us focus on the other experiments that showed greater differences.

The following experiments were based on the important role of the amygdala in the context of the visual system. As prior studies with SM showed, the absence of an amygdala which is suggested to act as a mechanism for emotional saliency, lead to her looking at some aspects of the scene differently. While SM might look at the face rapidly as normals do, we suspected that she might not look at features in the face in the same way. Prior studies with SM showed that when asked to judge the emotions in images she fails to look at the eyes. This is especially important when asked to identify fearful emotions, as shown with controls. SM’s inability to look at the eyes is not due to impairment in her eyes or systems that guide the eyes physically, but due to her lack of desire to look at these in order to judge emotions. Simply put – SM doesn’t find the eyes as salient given the task. Figure 93 shows a diagram of the visual system with the amygdala as its center module that guides the eyes towards emotionally relevant cues. Clearly the amygdala is highly important for guiding the attention in the visual system.

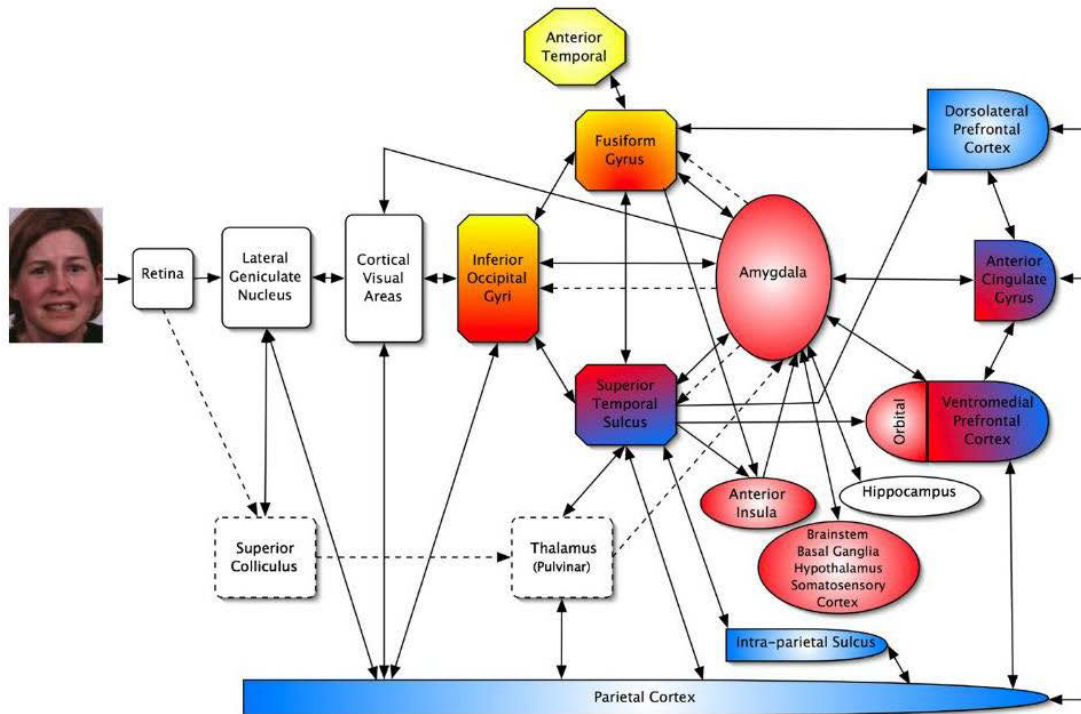


Figure 93 – Face perception and attention systems

The three rectangles with beveled edges indicate the core system for face perception (Haxby et al., 2000). Areas shaded in yellow represent regions involved in processing identity and associated semantic information, areas in red represent regions involved in emotion analysis (R Adolphs, 2002), and those in blue reflect the fronto-parietal cortical network involved in spatial attention (Hopfinger, Buonocore, & Mangun, 2000). Solid lines indicate cortical pathways and dashed lines represent the subcortical route for rapid and/or coarse emotional expression processing. This model is highly simplified and excludes many neural areas and connections. In addition, processing is not strictly hierarchical (i.e., from left to right) but involves multiple feedback connections (Bullier, 2001). The face displayed is from the database collected by (Gur et al., 2002). Image adapted from “Are you always on my mind? A review of how face perception and attention interact” (Palermo & Rhodes, 2007).

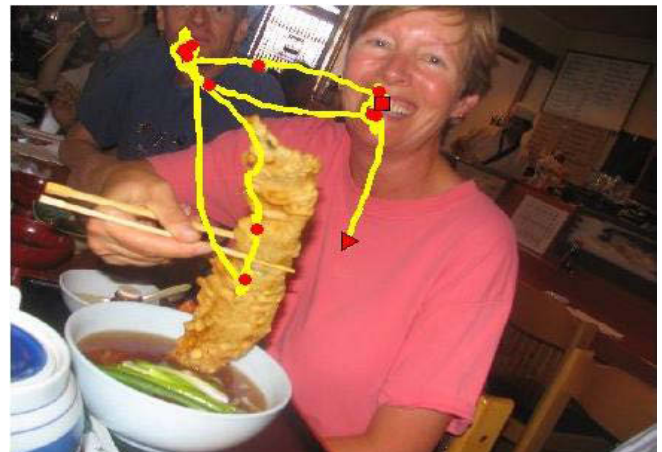
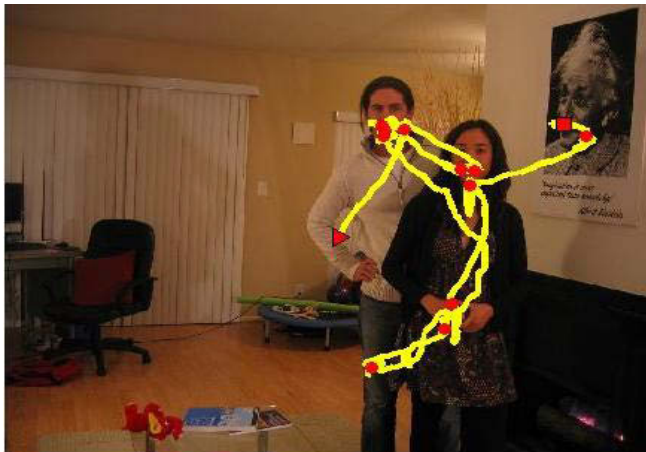
However, these images were put outside of natural context. SM was looking at Ekman faces in full-screen size, and was asked to identify the emotion independent of a scene. I was interested in seeing whether within the images I showed, SM would look separately at features of the faces regardless of the task (which remained to judge how interesting the images are (“free viewing”). I was, therefore, looking for differences in the viewing based on two criteria – full-size faces versus faces in a natural context. Familiar faces versus unfamiliar faces.

Methods

We collected a dataset of faces of people who interacted with SM in the course of the last 10 years from our lab and other labs, including pictures of SM herself, and embedded these in the overall task of block 1 in my experiment. We included very many images of people with whom SM as never interacted before. The images we accumulated were either shown in a full-size context, similar to the way by which images are commonly shown to SM in various emotional judgment tasks, or in natural scene context, as she would typically encounter them in her day-to-day activity.

Figure 94 shows images from both categories (close-ups and natural scenes; familiar and unfamiliar subjects) with SM's scanpaths.

SM was eye-tracked in a similar fashion to that reported for subjects in the prior eye-tracking experiments, while her responses and verbal interaction with me was recorded using an MP3 recorder as well.



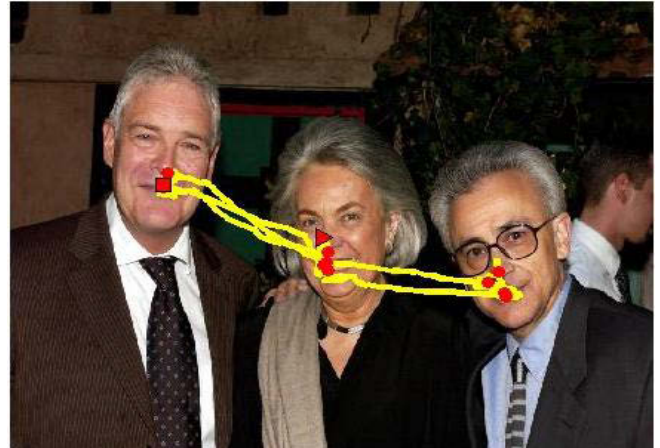


Figure 94 - Scanpaths of subject SM in natural scenes images with people

The people in the images were either familiar to SM (as are the Damasio family shown in the lower panel) or unfamiliar (as is the Asian girl in the top left panel).

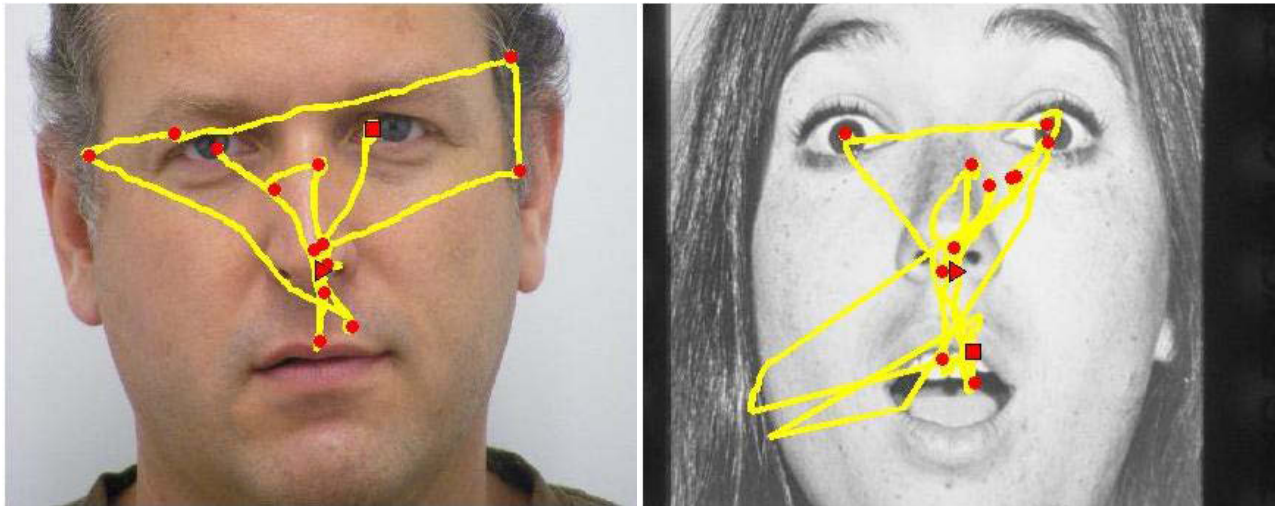


Figure 95 - Scanpaths of subject SM in full-sized faces images

Images either did or did not contain emotion context (left - neutral; right - surprised).

Results

While the results for the full-size images – which either do or do not carry emotional context – remain similar to these of controls (that is SM looks at the eyes and mouth equally, in a similar order, and with similar timing), the results for faces in natural scenes are very different.

SM seems to fail to look at the eyes in a social context repeatedly. In most images with faces, SM scanned the faces as did the controls, but was looking mainly at the mouth and not on the eyes. We analyzed (Figure 96) the proportions of fixations on the eyes compared to fixation on other regions of the faces (mouth, nose). Since mouth, eyes, and nose are very small in many of the images and are well within the margin of error of our eye-tracker we used the following method to quantify each fixation's location. We marked the center of each eye, the nose and the mouth, as well as the regions outside of the face (titled "out"). For each fixation we measured the distance to each of the locations and assigned that fixation to the bin corresponding to the closest region. That is, if a fixation is closer to the mouth than it is to the nose or eyes it will belong to the mouth category. SM spends 55% of her fixations on mouth regions, while she spends less than 10% of the other facial regions combined. Since she is looking mainly at faces, she still spends more time looking at people's mouth than she does exploring the environments ("out"), which is only 32%.

Importantly, these results (as was earlier mentioned) do not replicate when the faces are full-sized, and taken out of social context. This suggests that this method shows a different variant of SM's explorative behavior than the one tested in prior experiments with SM.

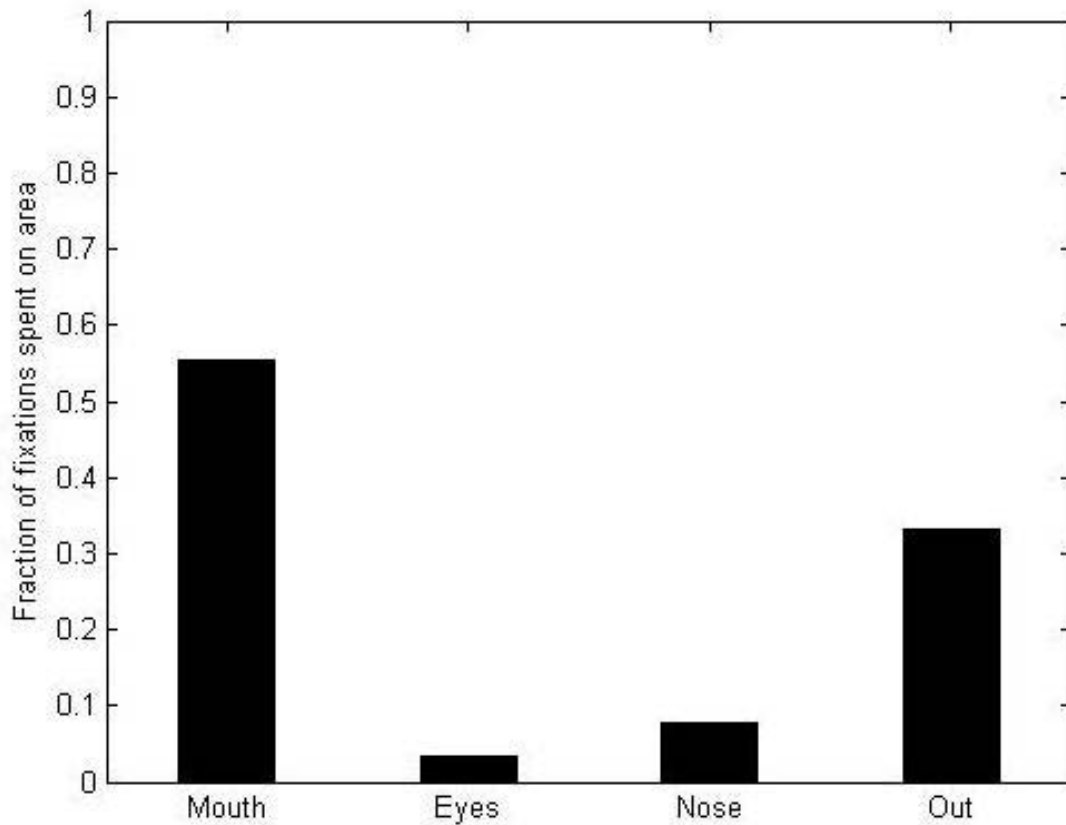


Figure 96 - Subject SM looks mainly at the mouth

Fraction of fixations spent on each of 4 regions in the image. “Out” corresponds to any region outside of the face, while the other 3 are manually defined by the center of the target (mouth, eyes, or nose). Looking only at mouth, for example, will give a fraction of 1 (100%). Since each fixation was placed in one of these four bins, the total of the fixations accumulates to 1. Mouth regions are looked at 5 times more than eyes and nose combined.

Testing for the effect of familiarity on the results shows no significant effect. SM looks at faces of familiar and unfamiliar people just as much and just as rapidly, and explores mainly the mouths of these two groups equally. While running a control study for the familiarity was not done due to the hardship of generating a unique set of images with people the subject is

familiar with for each individual and comparing those to others', we repeated the experiment with 12 controls in an undergraduate class at Caltech. These controls were looking at the same dataset, performing the same task. The results show that as expected, eyes are attended approximately 3 times more than nose (45% of fixations fall on eyes regions; 15% on nose), or mouth (10%), suggesting that indeed the bias to look at mouth, or "away from eyes" is unique to SM ($p < 10^{-6}$, Wilcoxon).

Experiment II – Attention allocation to faces carrying emotions

As I was genuinely fascinated by one of the main results reported by the Adolphs group back in 2005 showing that SM fails to look at the eyes when asked to identify fearful emotions, reaching a low performance in recognizing these emotions, I designed a follow-up experiment to this one.

First I will detail the results of the previous experiment. SM was asked to look at dozens of images containing the famous Ekman faces carrying emotions from 7 groups: Happy, Sad, Fearful, Disgusted, Neutral, Angry and Surprised. She was asked to identify the emotions in each image while being eye-tracked. The results show as abovementioned her failure to recognize fearful emotions, while looking mainly at the mouth in these images rather than the eyes, which was shown to be the relevant region for fear identification in controls. When asked directly by Ralph Adolphs to look at the eyes in a repeated block of the experiment her performance increased to that of controls, implying that the lack of amygdala leads SM to not desire to look at the eyes for the fearful emotions judgment, but when explicitly told to look at the eyes – replacing the role of the amygdala with a cognitive directive – she is able to fully perform the task just as well as controls. In a third block, when SM was asked to repeat the task

yet again, without any guidelines, she resorted back to her natural tendency to not look at the eyes, and indeed fell back to her low performance.

I was intrigued by one thing the most when I heard about this task and the results. This was SM's answer to why she does what she does and her explaining this behavior. I was interested in following this experiment a step further, and actually showing SM the results of her performance in the two blocks and trying to have her understand the decrease in the performance and see if she could learn to improve in the task and not fall back to her natural tendency as well as understand if she can shed light on this lack of desire to look at eyes. Essentially, as I put it when I spoke about the idea first with Lynn Paul: "Up until now we reported the methods and the results, and tried to interpret those ourselves. I want to repeat the methods and the results, but have SM try to explain the meaning of these to me herself. I want to know what really guides her attention in this particular task. If we write the Introduction/Methods/Results... I basically want to have SM write the discussion."

'Discussion'

SM did not write the discussion. I ended up having to write it myself. That said, there were few interesting milestones worth reporting on the way to figuring out why the lack of amygdala reduces SMs desire to look at the eyes when judging emotion scenes.

First, I wanted to "spice" the dataset to be used by adding to the standard Ekman faces a few new faces. I wanted to do that both because I suspected that maybe by now a well-studied subject is somewhat familiar with these faces and might already know the correct answers to some of the images just because of her experience with them. Second, I wanted to add faces of emotions depicted by SM herself. I figured that while SM is not necessarily good in being able

to identify emotions in the Ekman dataset, she should surely know one person's emotional expression the best – herself. She is the one person whose emotions she surely knows. I asked SM to pose for the camera in 10 different poses of emotion in each of the 7 categories. That's when the first intriguing anecdotal result came about. SM easily smiled during the 10 shots taken for happiness¹⁷, she showed a saddened face in 10 other images. Same good results for “neutral” and so on. The problem arose when I asked SM to mimic a fearful face. SM repeatedly failed to look fearful according to what I considered a fearful face. “This is not a fearful face,” I insisted as she smiled to me claiming that the smiling happiness on her face was in fact a fearful gaze. “I am doing my best,” she replied.

I asked Julien Dubois (who happened to pass by) to demonstrate a fearful face, and when he did I tried to use a different method. I asked SM to mimic his facial expression. “Just do the same face that Julien does,” I suggested. But SM kept being unable to do so. Eventually, I decided to give up on this and had SM do the face she claimed was fearful 10 times and used these images as the fearful faces for her experiment. Needless to say that in the follow-up rating of the images by controls for this study none of the faces SM claimed were fearful were identified as such by 10 subjects who looked at all the images and placed them in one of 7 emotional categories.

Being intrigued by SM's inability to not only identify faces, but even to demonstrate a fearful face herself, I went one step further. I sat next to SM and searched online with her the term “fearful faces” in Google images search. Dozens of images of fearful faces showed up on the

¹⁷ None of these images are included, as to keep SM's privacy.

screen. “Which of these images you identify as fearful?” I asked SM. I wanted to see how far her inability to identify this emotion goes, and to find some images that I could use for the experiment that would surely be identified by her as fearful. I needed at least a small number of images that I know she thinks are fearful to look at her eye-tracking when viewing these.

SM claimed none of the faces that showed up are in fact fearful. “They don’t look scared to me,” she said. I iterated the fact that the searched term was “fearful faces”, but SM insisted that none of the images looks fearful to her. Figure 97 shows 3 of the images that showed up and were not identified by SM as fearful.

I spent quite some time looking one by one at many images with SM trying to find images that she will find fearful. Eventually we found a single one that she claimed was scared (Figure 97).

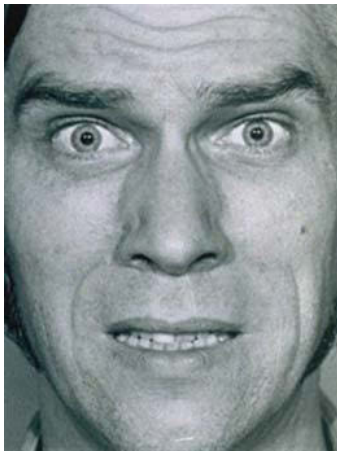




Figure 97 - Fearful faces from Google image search

Top. These three images were not identified by SM as fearful. For the 2 images on the right SM claimed the emotion depicted in them is surprise, and for the image on the left the emotion was anger.

Bottom. Image that SM identified as indeed fearful

Although I find all four images depicted in Figure 97 to be fearful I can in all honesty see that the level of fear and anxiety demonstrated in the image that SM found fearful is higher. That said, all four images were categorized as fearful with the control group so there was no evidence of levels in the ranking.

SM performed the task in a similar fashion to that by which she performed it in the prior study by Ralph Adolphs, but one with major change that I introduced. While in the original task for each image SM was given a 2-AFC between emotions to choose from, for each image I gave SM 7 alternatives. All the emotions were available for her to choose from, giving her a much wider range and making the task somewhat harder. This required a larger dataset to reach significance.

Results

SM indeed performed below average for the first task, and for the identification of fearful emotions in images of others and herself. While she was close to controls in identifying emotions such as Neutral, Disgust, Anger, Sad, and Happy, she performed poorly in identifying images which were fearful or surprised – repeatedly confusing one for the other (Figure 98). Looking at SM's eye-tracking results shows that she was looking equally at most areas, and did not show as clear a regional effect for any category. That is, unlike the previous results, she did not look at the mouth more than the eyes in this dataset, and exhibited a fairly higher number of fixations in the eyes even when judging fear. This can be due to training or practice, but prevented us from seeing a future significant improvement in the guided task where SM was directly told to fixate the eyes in the images.

That said, in the second block, where SM was directed to look at the eyes, and repeatedly reminded of that instruction when she failed to do so, the psychophysical results improved (Figure 99).

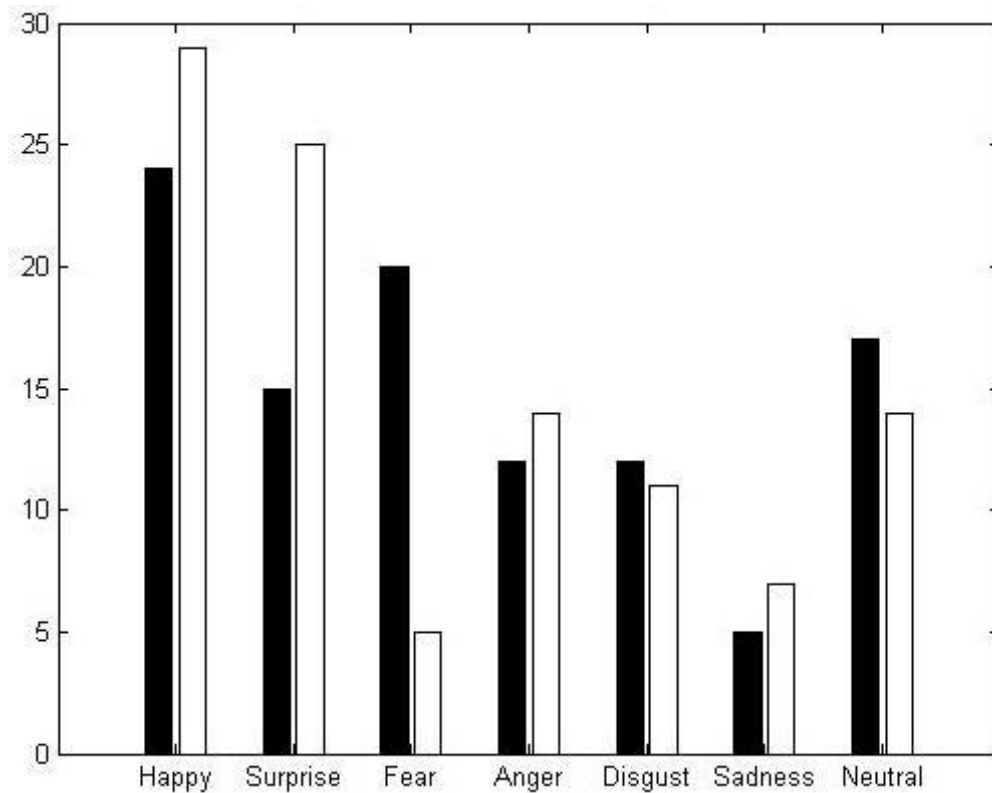


Figure 98 - SM judgment of emotions in images

White bars are SM's rating of images in each emotional category. Black bars are the ratings of 10 controls for the same images. Notice that SM was familiar with her own facial expressions that she depicted just minutes before the experiment, while the controls were all unfamiliar with any of the images. The highest difference in emotions' correct categorization is for fearful faces. In this task SM was free to look at the images as she wanted (undirected task).

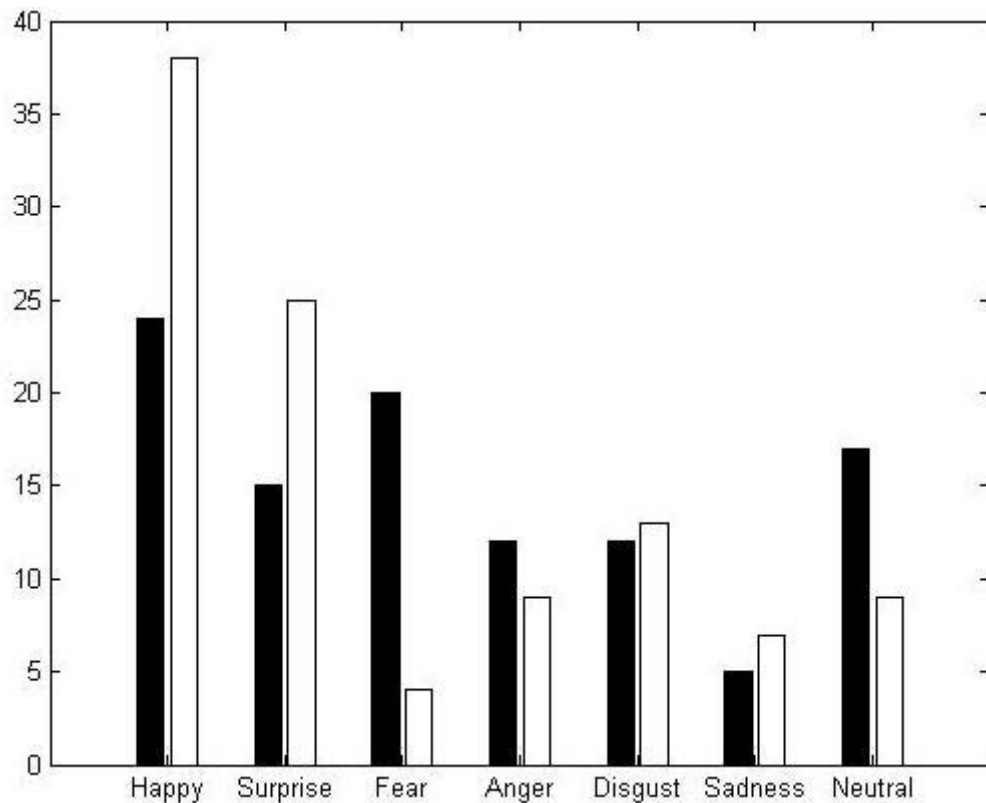


Figure 99 – SM judgment of emotions in images in a task guiding her to fixate the eyes

White bars are SM's rating of images in each emotional category. Black bars are the ratings of 10 controls with the same images.

The lack of improvement in the performance of SM in the second (guided) task did not allow us to fully pursue our planned experiment, as we could not claim that looking at the eyes yields an increased performance. While I did show SM the results of both tasks and confirmed with her that we feel that she typically would perform better in identifying fear when she attends to the eyes, the experiment could not progress as planned.

Interestingly, I asked SM to repeat the experiment slowly with me – following this stage – and stopped at each and every image where SM and I disagreed about the emotion, asking her to

explain to me (when she could) what made her categorize the images as this or that group. Most often SM showed a confusion between fearful faces, surprised, and angry ones.

Quantifying the mistake and showing which categories show the highest differences across the tasks, and in the follow up repetitions of the task, indeed confirmed that SM repeatedly confuses fear with surprise and anger (Figure 100). Based on results from valence and arousal studies positioning emotions on a grid this is in line with the overall relationship between emotions in healthy controls.

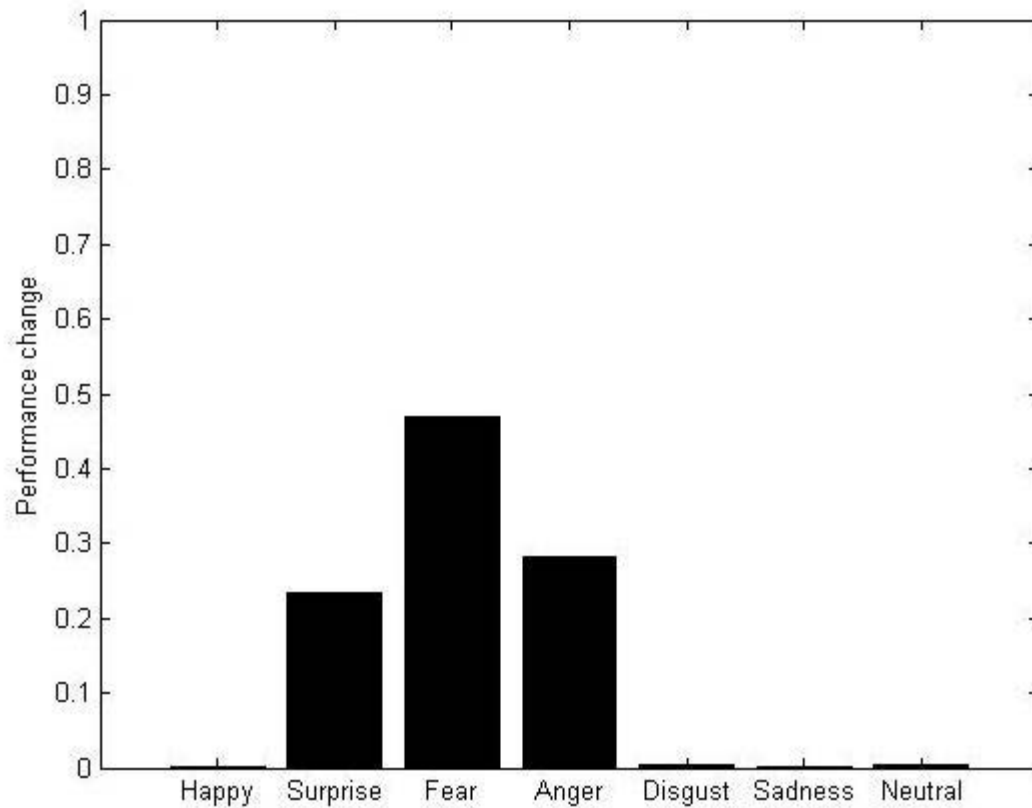


Figure 100 - Emotions performance change between the task

Bars represent the relative increase in performance of SM in identifying emotions after the feedback was given to her on her performance.

These results do not allow us to easily distinguish the increase in performance of SM based on learning and adaptation due to her understanding that she had made numerous mistakes in the first block, and her innate tendency to look at the mouth when she is asked to judge emotions.

An interesting analysis that will shed light on the innate tendency versus the acquired or learned-by-training results would be a comparison of these results (both psychophysics and eye-tracking) to the common images between the experiments that were conducted years apart. The fact that SM repeats similar behavior in the first block should suggest that she indeed desires to look at the mouth for fear identification and not just “avoid the eyes”.

Discussion

Finally, I went to ask SM to try and explain to me why she focuses on the mouth when asked to identify emotions rather than the eyes, and why she feels she has troubles identifying or recreating the feeling of fear.

SM has a very complex psychological background, and as part of the history and the cognitive need to provide a coherent story for herself regarding the lack of fear due to the amygdala lesion she has developed a concise story she tells herself wherein her complicated life circumstances had in fact led her to “overcome fear”. SM claims she in fact could experience fear beforehand, but learned to be brave and strong and hence not be afraid anymore. That is why she cannot depict fearful faces in front of the camera (“I do not know what fear is anymore”), and that is why she finds it hard to recognize it. While I believe that this is more of a personal compensation story that results from the actual lack of fear due to neurological reasons this is the desired “discussion” I hoped to have SM provide for me.

When asked if she would continue to look at the eyes in the future to identify emotions and use cues from other face regions than the face that I tried to convince her to, after showing to SM her scanpaths and convincing her that her lack of eyes is a unique effect that does not hold with controls, she was convinced that this indeed reflects a unique feature of her attention allocation.

However, she claimed that she finds it extremely hard to attend to other areas in the face when her [bottom-up] desire is to look at the mouth. “It’s as if I told you that you need to look at the left ear in order to know if someone is happy. You just won’t be able to do it because it won’t

make sense to you”¹⁸. What can I say... Attention is indeed a mechanism that is hard to override, even if we try to allocate resources to a competing region.

¹⁸ See appendix III for a transcript of the entire conversation with SM.

Parent of autism subject

One of my most interesting anecdotal results came from an autism subject who came to participate in an experiment with his mother. Since the subject was young and unlike most of my subjects had a rather higher level of autism (which made him borderline retarded) he was accompanied by his mom both as a guardian and as a mediator – helping me to get the instructions to him, while also taking care of him.

The rooms arrangement in Klab is such that while subjects participate in the experiment in our eye-tracking room I can sit in an adjacent room and follow the experiment through a screen replicating their monitor. This allows me to see problems that may arise in calibration for instance, and see the raw data as it is acquired (which commonly ends up being very useful). It allows me to tell if subjects have some difficulties in the performance or if subjects get tired or look outside of the screen, etc. – while at the same time not actually sitting with the subject in the testing room, giving him the liberty to perform the task at his own pace and at his own comfort level.

In this particular case I was sitting in the monitoring room with subject KA's mother, who was very interested in learning about the experiment and its results as she was caring for her child.

As I explained to her the results while the subject was running block 1 of the experiment (“free viewing” block, where subject sees images and is asked to rate them on a scale of 1 to 9 as to how interesting he finds each of them) she started guessing quietly the numeric ratings he would give to images as they were on the screen. “He would rate this one an eight”, she said just before her son indeed rated an image as very interesting – 8. “This one he will find boring, maybe a 3,” she said, while her son was looking on the numeric keypad for 3 in the other

room. As I listened to her predictions of his answers while looking at the screen with her I started noticing the level of accuracy in her predictions of his answer. They were all correct. She did not seem to mistake any of his numbers even though they did not seem trivial at all to me. Some of his answers were high, while the next were low. The spectrum was wide and she kept correctly guessing them. As I realized that this might be a fascinating view into the correlation between the parent of an autism child and his own ratings of images I immediately pulled a piece of paper and started frantically trying to record her answers, making sure that the following happened:

- a. She did not get to see his answers, so that the results would not be biased by mistakes of hers or by learning of his answers over time.
- b. She reported the answers as fast as she could – most often fractions of a second after the image appeared, to see how fast she can be in predicting his answers.

Sadly, since this was an ad-hoc unplanned experiment I had already missed approximately 60 images before I was able to start sampling the mother's answers in a somewhat organized fashion.

I had the mother and the son perform this “prediction-rating” task for an additional 3 other “free viewing” blocks, including block 7 (which includes only images with text), block 8 (which includes anomalous images), and block 5 (which includes images with somewhat disturbing content – the IAPS dataset (Lang, Bradley, & Cuthbert, 1995), including mutilated bodies, sexual scenes, explicit war images, etc.). Altogether we have results for all the images in block 1 as of image 57 (up to 200), the 187 answers for block 5, and the 80 x 2 answers for blocks 7 and 8.

Results

When I asked KA's mother why she's so good at predicting his ratings she answered that she "knows what he likes and what he feels passionate about." (It turns out that this was mostly animals.) He rated all the pictures with animals 8 or 9, and rated the images with masks, clowns, and stone faces 1-2 because it turned out that he had a "bad experience with clowns during his childhood and he doesn't like those at all".

Since KA's IQ was lower than our average high-functioning autism subjects, and his dependency on his mother was higher, we predicted that over time the mother "learns to become his translator" then it is interesting to quantify "how well is she translating him" in that sense.

We quantified the correlation between the answers separately for each block. For 2 of the blocks we didn't have enough samples (blocks 7 and 8 only have 80 pictures each), but for the other two blocks the correlation is: $r = 0.28$, $p = 0.0005$ for $n = 143$ for block 1 ("free viewing of images with faces"), and $r = 0.499$, $p < 10^{-13}$, $n = 187$ for block 5 ("free viewing" of images with faces and IAPS images; we lack 23 answers as these were instances where the mother was not quick enough to answer before he did). Calculating the Hamming distances for the two blocks showed that indeed the correlation was very high. Block 1 gave a distance of 0.657 and block 5 of 0.486. Block 7 reached 0.594 and block 8, 0.657.

Combine all the answers for all blocks and correlating these, gives an r value of 0.34, $p < 10^{-4}$, and $n = 477$ (187+143+74+73). The Hamming distance is 0.58. Figure 101 shows the results for all four blocks.

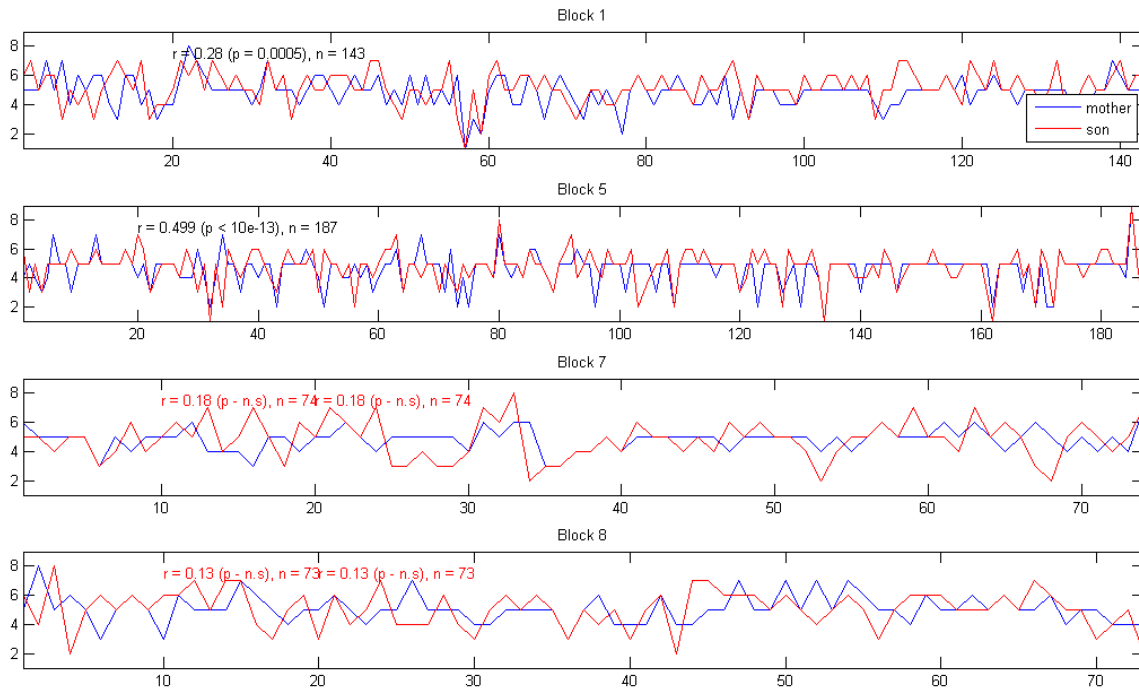


Figure 101 - Numerical ratings of images of mother and her autistic son

Each panel corresponds to a different block. Block 1 (upper panel) had 143 images, block 5 had 187 images, and blocks 7 and 8 had 80 images each. The red line are subject KA's answers, and the blue line are his mother's answers.

Conclusion

It is impossible to draw any clear conclusions from a dataset of $n = 1$. That said, it is suggestive of a very interesting tie between a mother of an autistic child and her son. In order to correctly control for this experiment one would need to test a few more parents of children with autism, (preferably with similar higher-level autism) and see the grouped results, while a similar trial should be used for parents of kids without the disorder, in order to see whether a simple strong connection between kids and their parents is leading to the high accuracy (parents who know their child well could better predict his answers to question), or if it has to do with autism. Naturally, parents of kids with a disorder such as autism have to care for him a great deal during the course of his life, so the intriguing result might purely be due to the close relationship between the two. A possible way to test this would be a simple interview conducted prior to the experiment with the mother and the son to rank the level of communication between the two and their closeness, which should predict higher correlation in the test results.

That said, since autism is very highly heritable, some of the parents are thought to process stimuli perhaps more similarly to autism subjects, and this might be a nice easy way to test this effect (as suggested in recent work “Distinct Face-Processing Strategies in Parents of Autistic Children” (R Adolphs, Spezio, Parlier, & Piven, 2008)).

Another interesting suggestion as a follow up to such a study would be to ask the mother to give her own ratings to these images rather than trying to predict her son's. While the first result sheds light on how well she knows and is able to predict his behavior, the second will actually suggest a correlation between the two subjects' personal ratings. This, however, would require a very large database of images and ratings in order to reach a significant difference from the

baseline – which is established in chapter 9 – showing that subjects tend to be highly correlated in their answers to images rating anyhow. A much higher correlation would have to be reached among the mother and son in order to argue for an even higher correlation than the one reached by any two subjects participating in such a task.

Finally, we might predict that in typical mother-son dyads, this correlation would decline as the son ages, but that the rate of decline is slower for mother-son dyads in which the son has autism. I wonder if the correlation between parent's prediction and the actual response would vary based on the degree of autism? Or on other features (such as IQ)? I would predict that with sons with lower IQ and greater behavioral impairment there would be a closer relationships with mothers, hence higher correlation in such a test.

Agensis of the Corpus Callosum

The final group which I tested for attention to social scenes using the experiment freely viewing images with faces and searching for faces and objects was the group of subjects with agensis of the corpus callosum (AgCC). These subjects are typically studied in the context of split brain type experiments and hemispheric transfer tasks. This is evidently due to the interest in the scientific community in the ways by which they overcome the lack of a corpus callosum – the brain's biggest bridge between hemispheres which is made of approximately 200,000 fibers.

However, as was mentioned in the introduction, typically AgCC subjects are known to have various social problems. We decided to try and test these subjects in a similar fashion to the ways in which we tested the autism subjects. Through a sequence of behavioral exams similar to these performed with the autism group (IQ test, stress tests, etc.) we monitored the behavior of these subjects.

Methods

Five AgCC subjects participated in the study (see Table 8 for details of the subjects).

ID	Sex	Age	IQ	Benton	STAI	
					State	Trait
1	F	40	100	47	22	41
2	F	32	97	45	45	44
3	M	32	88	50	68	66
4	F	52	84	41	34	54
5	F	18	84	47	27	27

Table 8 - Details on AgCC subjects

Subjects participated in the task in the same fashion as other controls and autism subjects.

Results

Figure 102 shows the proportions of first fixations on the face for the “free viewing” task in the standard analysis we conducted for all 5 subjects.

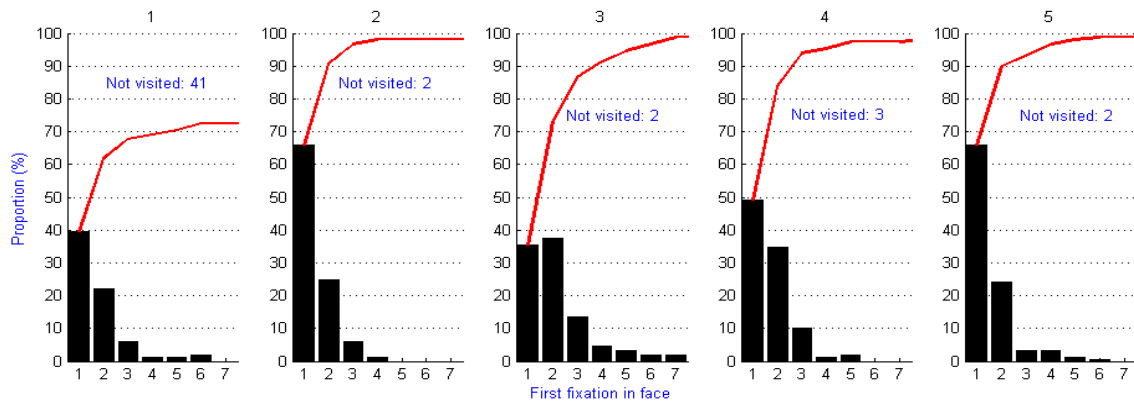


Figure 102 – Fraction of first fixation in face region for AgCC subjects in free-viewing

Each panel corresponds to the analysis of the proportion of first fixation in face for an individual subject. Black bars mark the proportion of first fixations; Red lines the accumulated sum of the proportions. For each subject we mark the total number of faces that the subject did not visit at all in the course of the entire experiment.

Subject 1 seemed to have been particularly low in her proportion of fixations as well as in her attention to faces *per se*. She did not attend to the faces in nearly a third of the images. Generally, the results of all AgCC subjects are comparable to those of autism subjects, well below the results of healthy controls. Comparing the results of AgCC subjects to those of controls or autism subjects revealed a dramatic decrease in proportions of first fixation in the eyes (Figure 103).

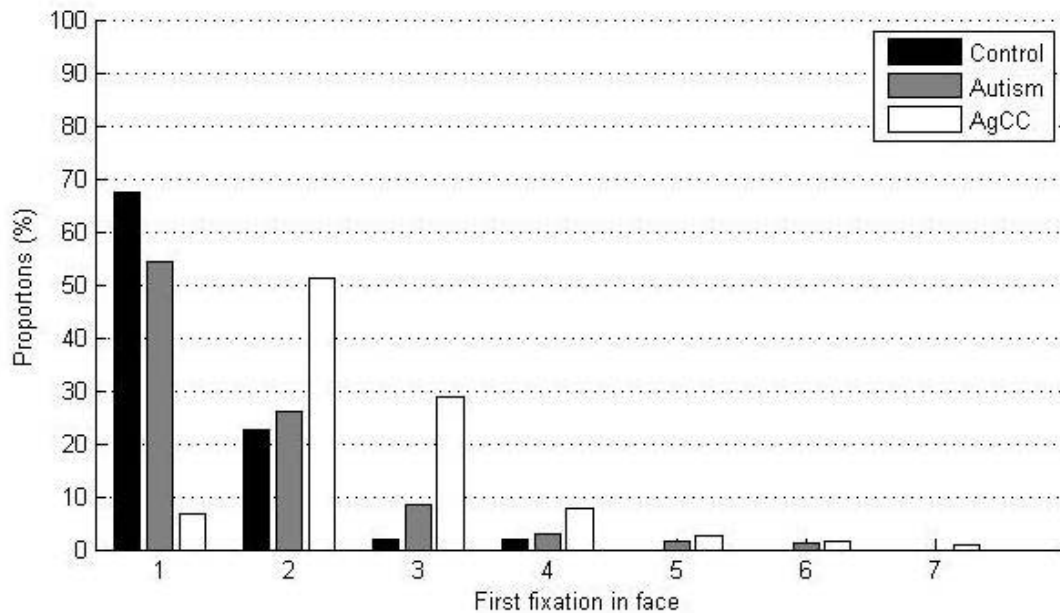


Figure 103 - AgCC subjects' proportion of first fixation on face region

Compared to controls (black) and autism (grey) subjects' proportions of first fixation on face regions, one can clearly see that AgCC subjects attend the faces mainly from their second fixation rather than the first.

AgCC subjects had viewed 90% of the faces only by their fourth fixation. AgCC subjects' viewing of faces in a "free viewing" task is different than that for both controls and autistics ($p = 0.07$ and $p = 0.01$, respectively, Wilcoxon).

For the "search task" the results were similar to those of controls and autistics (Figure 104).

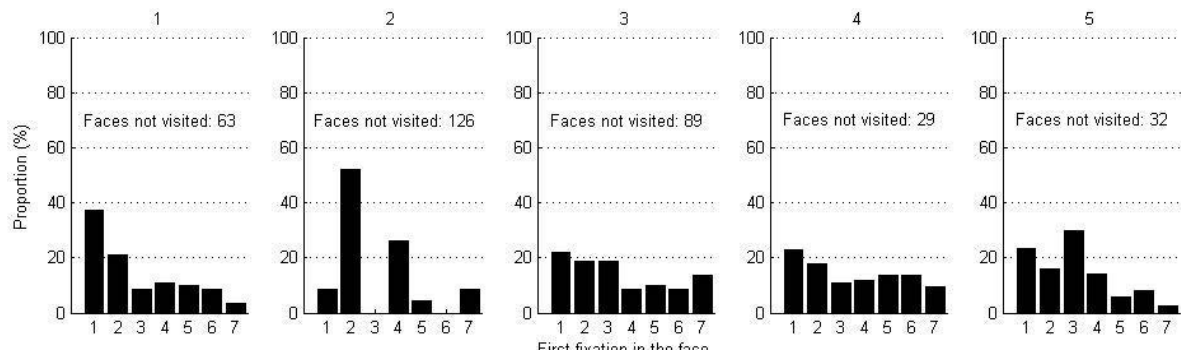


Figure 104 – Distribution of first fixation on the face for AgCC subjects in search task

Each panel represents the results of a single subject in the search task. Bars represent the proportion of first fixation on face regions. We mark for each subjects the number of faces not visited in the course of the task. Notice that subject 1 that had the lowest proportion of first fixations in the free viewing task and actually maintains a similar distribution for the search task, while subjects 2 and 3 – who had nearly perfect viewing of faces – drop significantly in this task. Subject 2 now nearly never visited faces at all.

While AgCC subjects show significant differences in the free viewing task, they seem to respond utterly the same as controls and autistics in the search task (Figure 105).

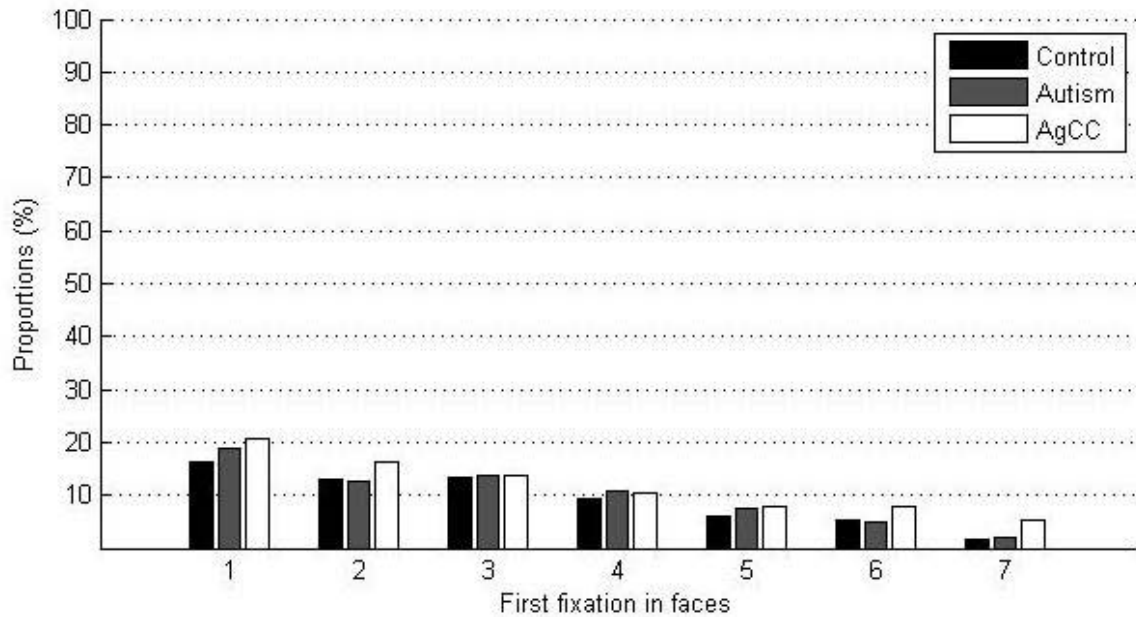


Figure 105 - Proportion of first fixation in face regions for the search task in AgCC

The distribution of first fixation on faces in the search task for the AgCC group look nearly identical to those of controls and autistics, suggesting that when directed where to look the AgCC group performs equally to other groups.

Discussion

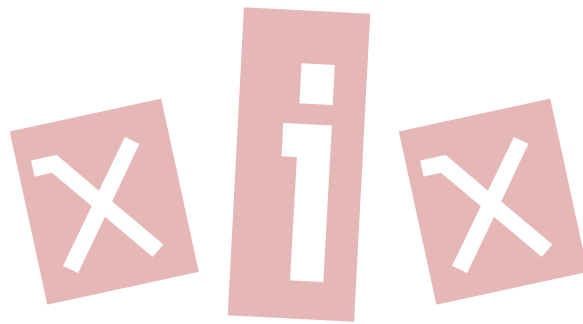
The results suggest an even lower level of attention to faces among subjects with agenesis of the corpus callosum. This supports claims regarding the social deficiencies among AgCC subjects, as they show lower interest in social scenes when opting to freely view scenes with people in them. The dramatically decreased results in the free viewing task, compared even to the already low attention allocated by autism subjects, hints toward a deficiency in social aptitude due to the lack of bridge between the hemispheres. Interestingly, the results of the search task, showing typical results that go hand in hand with these of control and autism groups show that AgCC subjects are in fact capable of looking at faces as much as any other group when directed to —evidence that their brains are capable of the task — yet they elect not to do so of their own free will.

While we could not identify a clear competing entity that captured AgCC subjects instead of the faces, the lack of interest remains evidence that the saliency mechanisms governing their attention allocation in still images are significantly different than those of controls. Further studies of social skills and their manifestation in the corpus callosum might shed light on the neurological reasoning behind this lack of interest.

In the competition context we can look at the information flowing from the retina to each area regarding the scene, and potentially through the fusiform face area too, as not reaching a critical mass to give rise to a saccade towards the face because of the lack of information transfer between the hemispheres that is crucial for the accumulation of information that yields a saccade towards the face. In the absence of the aggregated information from both hemispheres the brain is left to allocate attention based on other lower mechanisms that, in a

direct competition with information from both hemispheres, would have failed to result in a saccade, but now garner it. A method for testing this will be using fMRI or other imaging techniques to trace the flow of information, specifically through face regions, while the task is performed.

This task, and its significant result, can mainly be useful in supporting claims that AgCC subjects indeed lack some social skills or attention due to their unique neurological condition and as such should be treated with additional care when their social skills are tested. While typically the “awkwardness” of some AgCC subjects was not considered a direct result of their missing corpus callosum, experiments such as this and results such as the ones we show here could benefit the AgCC community by offering yet more evidence of the differences in perception of social occurrences due to their neurological condition.



Attention in Marketing

Attention is a prerequisite for all marketing efforts.

Sacharin

R¹⁹esearch that integrates findings from cognitive psychology, cognitive neuroscience, and marketing is in its infancy. Nevertheless, a few marketing researchers ventured into this brave new world that is expected to hold much potential for advertising research (Vakratsas & Ambler, 1999). Merging cognitive neuroscience with research on consumer behavior offers tremendous potential for growth in knowledge. This is especially so because “a great mismatch exists between the way consumers experience and think about their world and the methods marketers use to collect this information” (Zaltman, 2003).

Conveniently, neuroscience uses new technologies that make it possible to measure neurophysiological activity in order to study complex human behaviors (for a good overview of neuroimaging methods in terms of relevance to consumer behavior research see (Egidi, Nusbaum, & Cacioppo, 2007; Plassmann, Ambler, Braeutigam, & Kenning, 2007)). These tools carry the potential to override the methodological problems of the former approaches (Plassmann et al., 2007), although whether this will actually occur is yet to be seen. The hope is that these physiological measures will be used to augment, not replace, traditional research methods in order for marketing researchers to begin to more adequately validate and refute some of the long-debated theories of consumer behavior, and, by default, human behavior in general.

Within this broad research framework, of interest to the present chapter is one currently ignored, but perhaps crucial, aspect of consumer behavior: selective attention. A recent review

¹⁹ This chapter is partially based on: Milsoavljevic M, Cerf M., “What Matters is Attention not Intention: Insights from Computational Neuroscience of Vision”, *International Journal of Advertising*, Volume 27, Number 3, 2008

of how neuroscience can inform advertising (Plassmann et al., 2007) shows that no studies have so far examined the construct of attention. This is surprising, given that marketing researchers declared attention to be a prerequisite for all marketing efforts (see introductory quote). Thus, selective attention should be included within the neuroscience-marketing research framework.

Another omission in the emerging field that combines neuroscience and marketing is that insights from theoretical and computational neuroscience have yet to be introduced. Computational neuroscience combines what is known about the brain from neuroscience with the computing power available to simulate neuronal and psychological processes on a computer (Sejnowski, Koch, & Churchland, 1988). The goal of computational neuroscience is to develop algorithms that can simulate on a computer the ways by which the brain functions when we perform tasks (E. Smith & Kosslyn, 2007). Although many computational models of memory, attention, learning, and decision-making have been introduced, the present chapter focuses on perhaps the most developed, computational models of visual attention.

Thus, there are two objectives of the current chapter. The first objective is to bring into the spotlight the construct of attention. The second is to introduce computational neuroscience of visual attention to the marketing field and discuss its utility for understanding deployment of attention in the advertising context.

The construct of attention in advertising context

Due to a plethora of communication channels, consumers are faced with an overabundance of information – a typical consumer is exposed to several hundreds (even thousands) of marketing messages daily. Not all of this information can be processed because of the limited capacity of the brain, known as the attentional bottleneck. In recognition of this cluttered

environment, some researchers have declared that we are living in the attention economy, with attention being the scarce resource (Davenport & Beck, 2001).

Not surprisingly, everyone involved with marketing knows the importance of getting consumers' attention. In advertising, the importance of attention is evidenced by attention's prominent position in many advertising models. Originating in 1898, the first formal advertising model, AIDA: Attention → Interest → Desire → Action, positioned attention as the first step that people go through when exposed to advertising and before making a purchase (Vakratsas and Ambler, 1999). Furthermore, most hierarchy-of-effects models suggest that attention is a necessary step before higher-level processes can occur.

Since the importance of attention to marketers is factual, it is surprising that marketing studies of attention are rare (Rosbergen, Pieters, & Wedel, 1997). The little space that attention does receive is devoted to describing how attention is measured, while little emphasis is placed on any conceptual discussion, as is described next.

The concept of attention

Even today, most marketing researchers' understanding of attention is limited to William James' view dating back to 1890: "Every one knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought." A single exception is a recent stream of research of Pieters and Wedel (Pieters & Wedel, 2004, 2007) who suggest that attention is a much more complex phenomenon than is currently studied in marketing.

Pieters and Wedel (Pieters & Wedel, 2004) introduced two determinants (found in psychology and neuroscience) of attention to advertising. The two determinants are: (1) bottom-up and (2) top-down attention.

Bottom-up attention is a rapid, automatic form of selective attention that depends on intrinsic properties of the input, such as its color or intensity (C Koch, 2004). It is also known as saliency-based attention, indicating that the more salient an object is, the higher the probability of it being noticed. Top-down attention is a volitional, focal, task-dependent mechanism often compared to a spotlight that enhances processing of the selected item (C Koch, 2004).

In the present work, bottom-up processes are referred to as preattention, while top-down processes are referred to as focal attention. Thus, attention is viewed here as a two-step process, consisting of preattention and focal attention, although this is not necessarily a sequential process, as top-down attention can sometimes moderate the bottom-up processes (Cerf, Frady et al., 2008).

Pieters and Wedel (Pieters & Wedel, 2007) made important first steps toward improving our understanding of how top-down factors (i.e., consumers' goals: "memorize the ads", "collect brand information", "evaluate product", etc.) may influence attentional deployment within magazine ads. However, it is also important to spark research on the effects of fast, bottom-up attention. This type of research is virtually non-existent in the marketing literature. As the work of Pieters and Wedel is focused on print advertising, the same context will be used in the remainder of the current chapter to demonstrate the importance of research on bottom-up attention.

The research on bottom-up attention may be especially important since, as with the other types of media, clutter is an imposing problem for magazine advertising. For example, a 318-page issue of *Glamour* magazine contains 195 pages of advertisements and 123 pages of editorial content (Clow & Baack, 2006). Faced with such amount of clutter, consumers often have a singular goal: “to avoid advertising”. Even Pieters and Wedel (Pieters & Wedel, 2007, p. 224) highlight that “processing goals may have a lower likelihood of surfacing during the few seconds that consumers typically spend on ads during self-paced exposure” and that “competitive clutter may favor reflexive control and hinder systematic goal control”. Also, they state that attention to ad objects is very low during free viewing of magazines because object salience, or bottom-up driven attention, primarily determines attention allocation during free viewing (Janiszewski, 1993; Pieters & Wedel, 2007). Thus, on many, although not all, occasions, bottom-up attention may be as close as advertisers can get to consumers before the top-down goal of “ignore advertising” kicks in.

An earlier study of Pieters and Wedel (Pieters & Wedel, 2004) serves well to further argument this point. They instructed more than 3,600 consumers to freely browse through magazines, and used eye-tracking to measure where consumers directed their gaze. The experimental magazines included 1,363 print ads. They found that on average, 95.7% of participants fixated at least once at ads, but that the lowest scoring ad was skipped by 39% of the participants. Further, people spent on average 1.73 seconds with each ad, ranging from .037 seconds to 5.30 seconds.

This study shows that even when people are seated in a laboratory and asked to look at a magazine, the time spent with ads is very low. It can be assumed that time spent with ads and

the number of ads that people look at are even lower during natural magazine browsing when people's attention is also consumed by other factors in their environment.

Thus, the construct of attention in advertising should be studied based on the two-component framework, consisting of bottom-up and top-down attention. The biggest challenge in such undertaking is that, while marketers recently began using improved measures of focal attention, measuring preattention is still challenging.

Measurement of attention

Perhaps the most common method of measuring attention to advertising is by using self-reported memory measures ("To which extent did you pay attention to this ad?"). However, memory measures are poor indicators of what consumers pay attention to (Rosbergen et al., 1997) for at least two reasons:

First, attention is known to precede awareness. Thus, it is possible that a stimulus was attended to, but has not reached the awareness stage, thus making it impossible for individuals to have it in their memory or to report it.

Second, even if a stimulus was attended to, people are known to forget most of the stimuli they process.

A somewhat improved, although less frequently used, method of measuring attention in marketing is eye-tracking, where eye-movements are recorded to indicate individuals' attentional patterns. The main weakness of eye-tracking, as currently used in advertising, is that findings are "rather superficial" (Rayner, Rotello, Stewart, Keir, & Duffy, 2001). For example,

the size of the advertisement has been found to influence participants' looking times, which was pointed out by psychophysicists over a century ago (Tatler et al., 2005).

Thus, although some progress has been made in studying and measuring top-down attention, these methods do not account for bottom-up attention. What may prove to give life to research on bottom-up attention is a branch of neuroscience known as computational neuroscience. To initiate this stream of research within marketing, computational neuroscience of visual attention is introduced next.

Computational neuroscience of visual attention

The goal of computational neuroscience is to relate the data of the nervous system to algorithms employed by the brain to carry out higher-level human behaviors such as attention, learning, memory, emotions, and decision-making. These aim to create a computer simulation of real human behavior inspired by biological systems in the brain.

Computational brain modeling attempts to produce either (1) realistic brain models or (2) simplified compact brain models. A realistic brain model is a large-scale simulation that goes to the level of a single cell (Sejnowski et al., 1988). Since the model becomes much more realistic at the cellular level, it becomes less helpful in understanding its function at the nervous system level. Further, realistic simulation is computation-intensive, meaning that it requires a substantial computer power.

On the other hand, simplified brain models are networks of brain cells known as “neural networks,” which capture important principles of the functionality of a system (Sejnowski et al., 1988). Most importantly, neural networks are being used as models for psychological

phenomena such as attention, emotions, and decision-making. Of those, perhaps the most realistic and profound computational models are those which simulate visual attention.

Vision means “finding out what is where” (Figure 106) and computational modeling attempts to provide algorithms which successfully locate and identify informative objects in a visual scene.

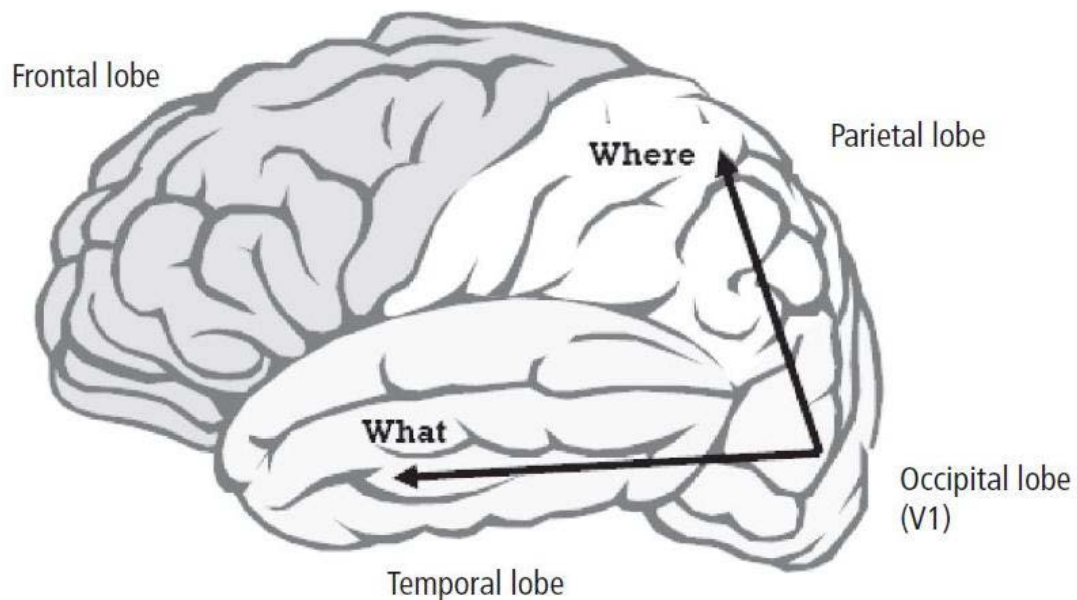


Figure 106 - Visual processing pathways

From (E. Smith & Kosslyn, 2007)

Specifically, two separate cortical routes are involved in vision, giving rise to two streams of visual information (for a detailed review see (C Koch, 2004). Spatial deployment of attention (“where”) is known as the dorsal pathway. It proceeds from the primary visual cortex (V1) in the occipital lobe, through the posterior parietal cortex, and to the dorsolateral prefrontal

cortex. Object recognition (“what”) happens via the ventral pathway, which involves V1, inferotemporal cortex, and the ventrolateral prefrontal cortex.

As mentioned before, attention is not given to all visual input. Since our visual environment is cluttered, attention serves as a processing bottleneck allowing only a selected part of sensory input to reach visual awareness. This process depends on the two previously mentioned mechanisms: bottom-up and top-down attention. This two-component framework of visual attention was introduced by Treisman and Gelade (Treisman & Gelade, 1980) and gave basis for development of computational models of visual attention. Guided by the idea that a visual scene is initially analyzed based on the physical properties of objects in the scene, the first neurally plausible computational algorithm of bottom-up attention was developed by Koch and Ullman (C. Koch & Ullman, 1985), and later extended and implemented by Itti, Koch, and Niebur (Itti et al., 1998). The model is briefly introduced next.

The model of bottom-up attention and saliency of Itti, Koch, and Niebur

The model’s flow (a rough outline of the model’s architecture is depicted in Figure 107) begins by analyzing physical characteristics of objects in a given visual image. It analyzes color, intensity, and orientation of objects and sorts these into three conspicuity maps, which are gray-scale maps where brighter areas represent more salient locations while darker ones are less salient. One conspicuity map is created for each of the three characteristics: color, intensity, and orientation. For example, an object may be identified as highly salient because of its color (the object’s location would be represented as a brighter area on the color conspicuity map),

while another object may be deemed salient due to its intensity (the objects' location would be brighter on the intensity conspicuity map).

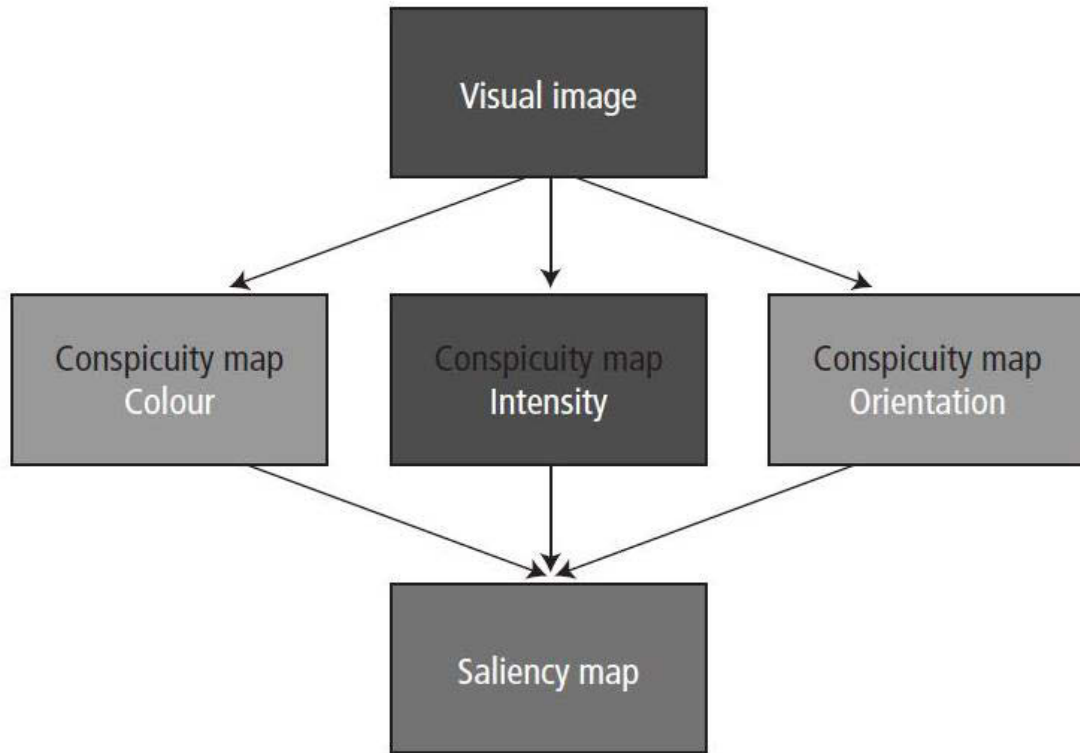


Figure 107 - Computational model architecture

The values from the three conspicuity maps are summed up into a saliency map, which is a two-dimensional topographic map that represents the saliency at every location in the visual image (Itti et al., 1998). The most salient locations are potential targets for visual attention (Schall & Thompson, 1999). The most salient location is identified first. Then, this location is inhibited in a biologically motivated fashion and the next most salient location is determined, and so on. In this manner attentional scanpaths are created for a given visual image (Itti, 2004).

The model of bottom-up attention and saliency discussed here is a neurally based model, i.e., it mimics human performance in a manner that is inspired by biological circuitry. Currently, it

is probably the most widely used model of bottom-up visual attention (Cerf et al., 2007). It has been validated for the past decade on a number of classical visual search experiments and was found to be “consistent with observations in humans” (Duchowski, 2007). Recently, the model has been tested and improved in a number of contexts, including the presence of (1) motion, (2) faces, and (3) text. These are briefly discussed next.

First, while the saliency model was initially implemented on static images, it could easily be scaled to motion — taking frame-by-frame video data and analyzing those as a single image. For example, recent additions to the model include flicker channels that allow for attention-allocation to rapidly changing content that was shown to attract human fixations and was found to be consistent with data from eye-tracking (Cerf, Frady et al., 2008).

Second, it was recently shown (Cerf et al., 2007) that adding a face channel — taken from existing face detector algorithms — can enhance the predictions of the model when it comes to telling what human observers are looking at in an image. Also, and relevant to marketing research, when observers are looking at faces in an image this can increase the total amount of time spent viewing the image.

Finally, since we are exposed to text very often in our lives, it was shown in a recent work of Cerf, Frady and Koch (Cerf, Frady et al., 2008) that inclusion of a text channel in the saliency model further increases the ability of the model to predict what people are looking at. The performance of the model has been compared to the eye-tracking data collected from people exposed to a number of different images, and was found to be comparable.

It is important to note that the model is consistently found to perform comparable to results obtained from eye-tracking. The cost-benefit properties of such findings are the following: no

expensive eye-tracking equipment is necessary, participants are not needed since the model simulates universal human bottom-up attention allocation, and significant time efficiency can be achieved as the algorithm can provide real-time analysis.

As this brief review of the current research on bottom-up attention shows, there has been much progress in modeling the process. It is important to remember that this sensory input is sometimes modulated by top-down, person- and task-dependent input (Cerf, Frady et al., 2008). Computational models that include top-down cues are currently being developed by several research groups but are not as well developed as the models of bottom-up attention (Cerf et al., 2007; A Oliva & Torralba, 2007).

The following section discusses how computational modeling of visual attention may benefit both marketing theory and practice.

Potential contributions of computational neuroscience of visual attention to marketing research

Using computational modeling of bottom-up visual attention in marketing studies has the potential of making significant (1) theoretical, (2) empirical, and (3) substantive contributions.

First, the conceptual understanding of attention will be enhanced by investigating factors that determine its key component – preattention. A better understanding of how to manipulate preattention will enable researchers to study its consequences, such as attitudes, intentions, and/or choices. The effects of preattentive processing on attitudes toward the ad and brand were mentioned in the context of mere exposure effects when Janiszewski (Janiszewski, 1998) highlighted the importance of investigating whether “preattentive processes [are] instrumental

in the formation of affective responses and, if so, how these preattentive processes operate.” About half a dozen marketing studies investigated the role of preattention within the mere exposure phenomenon but no clear conclusions have been reached (Janiszewski, 1993; Shapiro, MacInnis, & Heckler, 1997; Shapiro, MacInnis, Heckler, & Perez, 1999)

Second, although focal attention is being measured by somewhat improved methods of eye-tracking, the measurement of preattention is still challenging. Yoo (Yoo, 2005) highlights this by stating that one of the most important emerging issues in the study of preattention to advertising is “empirically detecting the existence of preattentive processing.” This can finally be addressed by introducing computational modeling of bottom-up attention, which *a priori* identifies objects that are likely to be preattentively processed by a viewer.

Finally, potential applications of computational modeling in the domain of advertising pretesting and evaluation are abundant. As an example, the following section demonstrates the utility of computational modeling of visual attention in the context of print advertising.

Application of computational modeling of visual attention to evaluation of print advertisements

As already suggested, many people purposely avoid looking at ads or look at the ads briefly. In such an environment, advertisers should be (and perhaps already are) trying to ensure that the key information in the ads is at least quickly, preattentively processed by consumers. This makes sense, since it is known in psychology and neuroscience that automatic, preattentive processing is very rapid, occurring within less than one second of exposure to a visual scene (Quiroga, Mukamel, Isham, Malach, & Fried, 2008). The previously described computational

model of visual attention (Itti & Koch, 2000) offers a tool which can be used in design of print ads to ensure that key elements of an ad are salient, and, thus, more likely to be at least preattentively processed by viewers during these brief exposures.

To illustrate this process, two magazine ads from the Procter & Gamble's Tide campaign, which won the 2007 Clio Award (available at www.clioawards.com/winners), are evaluated using a numerical computing software program – Matlab and the saliency algorithm of Itti, Koch, and Niebur (Itti & Koch, 2000) available at ilab.usc.edu.

Figure 108 shows the computational modeling output for the first ad (original images are in color, but are shown here in grayscale). The image of the ad is used as a visual input and is decomposed into three conspicuity maps (one for each: color, intensity, and orientation), which are then summed into a saliency map. The saliency map shows conspicuous ad objects, where the brighter the location the more noticeable the object is. Based on the saliency map, and as shown in the “Attention Guidance Map” in Figure 108, the most salient locations (circles) and the order in which attention shifts (lines and arrows) are identified. Also, the time required for each shift of attention is calculated by the program. In this example, the analysis simulates what an individual would preattentively process during the first half second of exposure to the ad.

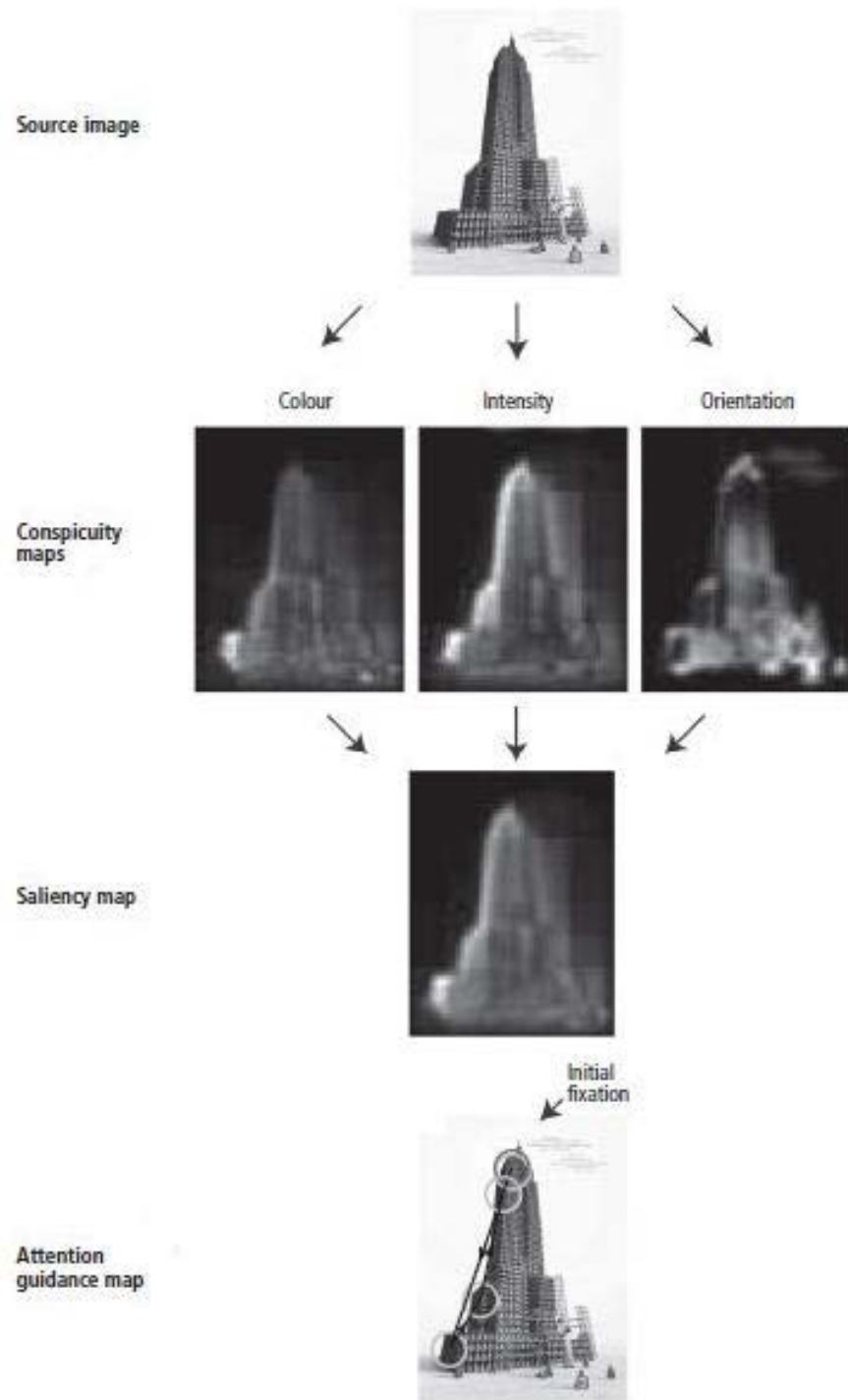


Figure 108 - Bottom-up attention to ad 1

The analysis shows that the most salient location is the left, dark side of the building (high intensity shown on the intensity conspicuity map), which provides no information about the product or the brand. Perhaps, a viewer may spend up to half a second on this shaded area before turning the page without receiving any useful communication about the brand or the product. Since many people are not likely to consciously pay attention to the ad, and computational modeling shows that the ad is not likely to be preattentively processed either, there is no reason to expect any positive advertising effects in this example.

The second ad comes from the same campaign and the output of computational modeling of this ad is shown in Figure 109. Once again, color, intensity, and orientation of objects in the ad are presented in the conspicuity maps, which are then summed into the saliency map.

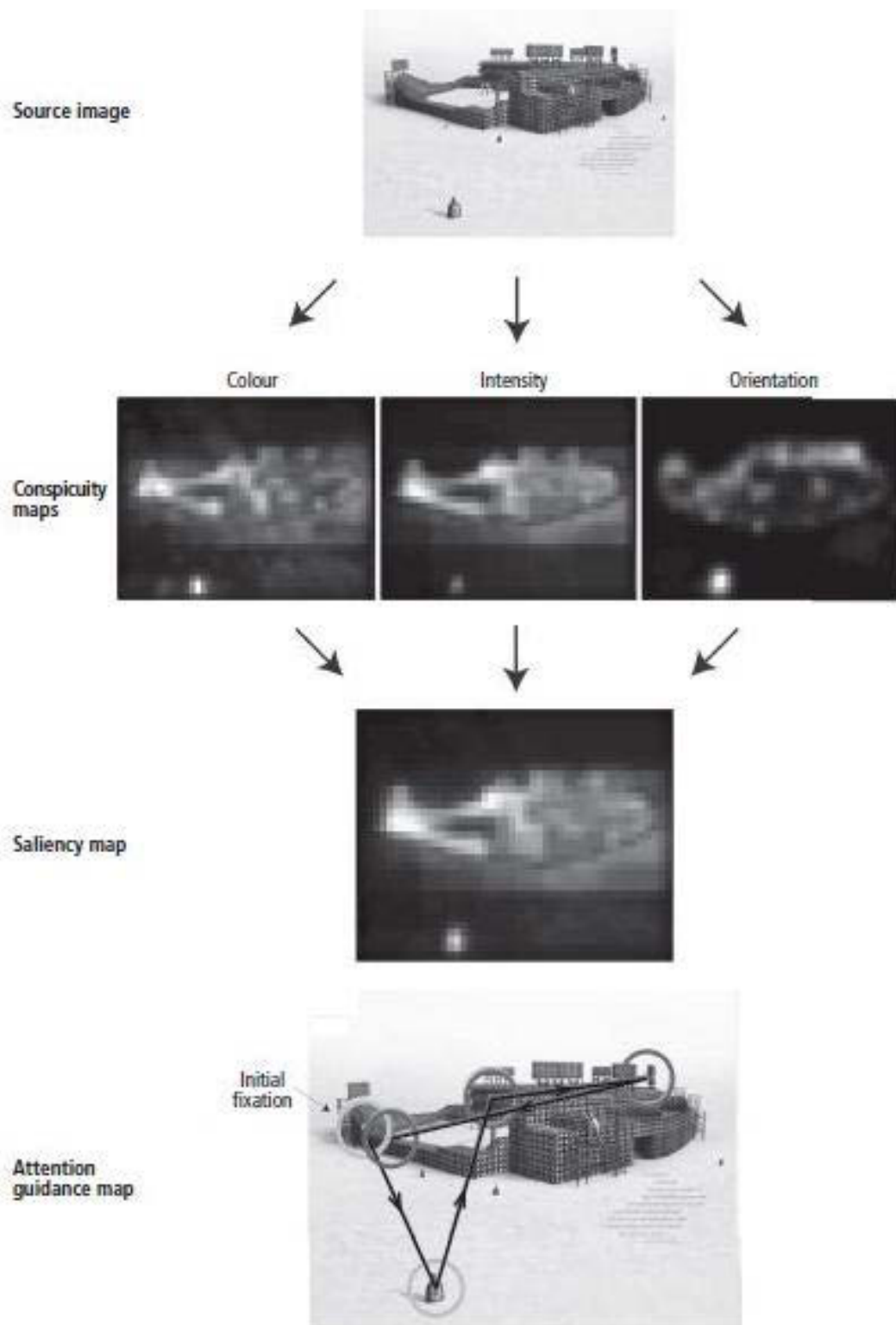


Figure 109 - Bottom-up attention to ad 2

In this case, the product packaging at the bottom of the image (circle on the bottom of the page) is the second most salient object (after the left corner of the stadium), and even the third most salient location is within the product bottle (the label is salient due to its color and then the top of the label becomes salient because of its orientation, as shown in respective conspicuity maps). Thus, even if a viewer consciously ignores the ad and spends as little as half a second on it before flipping the page, the brand name and product bottle are likely to be at least preattentively processed, opening a possibility for positive advertising effects. One can even argue, although much research is needed to prove this hypothesis, that a more relevant top-down goal, for example, “I am out of laundry detergent and need to buy some,” may at this point override “avoid advertising” goal, thus resulting in focal attention to the ad.

Once again, it is important to note that a major advantage of using computational model of bottom-up attention described here, if additional studies demonstrate its validity for advertising research, is that it does not require recruiting participants or employing time- and cost-intensive eye-tracking methodology.

Conclusions

The purpose of the current chapter was twofold: (1) to highlight the importance of studying attention within the emerging research paradigm that combines marketing and neuroscience, and (2) to introduce the field of computational neuroscience to the marketing discipline.

First, it was argued that attention currently receives virtually no space in the neuroscience-marketing literature, even though attention is a necessary step for all other marketing efforts. More so, it was emphasized that studying attention as a two-component construct consisting of a combination of bottom-up and top-down processes provides a useful framework for increasing understanding of the ways by which attention operates. In such a quest, the context of consumer behavior proves to be a very relevant and natural way to study and better understand attentional processes.

Second, using computational modeling of visual attention to simulate early attention on a computer offers much potential for improving conceptual understanding and methods of measuring preattention, as well as a host of opportunities for application in the field.

In summary, the current chapter identified a gap in the marketing and, more specifically, advertising literature, and thus has opened numerous research possibilities, some of which are briefly discussed next.

Directions for future research

Future research should uncover how preattention operates; for example, which physical characteristics of objects result in preattention, and when they are effective? Once a better understanding of preattention is achieved, the studies that assess the relationship between

preattention and focal attention, as well as between preattention and attitudes, emotions, and decision-making should, follow. This will enhance our understanding of the concept of attention, its antecedents, and its consequences.

The present chapter demonstrated the utility of computational modeling of visual attention in a specific advertising context, magazine advertising. Next, some recent preliminary studies that have yet to receive attention of marketing researchers are mentioned.

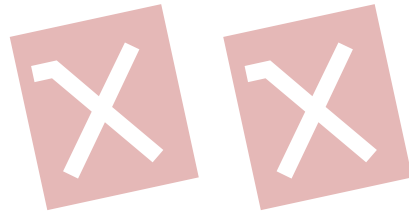
First, in the work of Torralba and Oliva (A Oliva & Torralba, 2007) it was shown that people can quickly identify the “gist” of the image, which can bias attention allocation. For example, attention is allocated faster to an image of pedestrians (expected to be walking on the ground) when they are shown at the bottom of the image than to an image of pedestrians located at the top of the image (A Oliva & Torralba, 2007). In the context of magazine advertising, for example, if we want to emphasize the shoes of the model in an ad, they should probably be placed in the bottom margin of the page, thus increasing the probability that the object will be noticed. Future research in advertising context is needed to test to which extent attention allocation depends on such general properties of a scene.

Second, the role of attention allocation by bottom-up driven saliency models has been studied in the context of video-gaming. Peters and Itti (Peters & Itti, 2007) recently showed that some bottom-up driven attention mechanisms may not only govern ad viewing, but even computer-game playing. Future research in this context is warranted given current escalation of in-game advertising.

Finally, in a preliminary study, the computational model of bottom-up attention and saliency was used to design banner ads on a website in order to make them more or less salient. In an

experiment where all other factors were controlled and only the saliency of the banner ad was manipulated, consumers' attitudes toward the banner ad were progressively enhanced as people spent increasingly more time on the website in the condition where the banner ad was designed to be salient, while no such change was observed for nonsalient banners.

It is important to note that, as argued earlier, recall and recognition rates for the target ad were very low (less than 20%) even in this extreme condition where most people spent 2-3 minutes exposed to the ad on various webpages. This provides further support for the argument that people often purposely avoid looking at the ads, and once again points to the importance of studying the confluence of bottom-up and top-down attention. The hope is that the current chapter will motivate both marketing academicians and practitioners to better understand the construct of attention and to perhaps do so by utilizing the knowledge and tools from cognitive and computational neuroscience.



Visual Projection of Thoughts from Single Neurons in Humans

We are pessimistic because life seems like a very bad, very screwed-up film.

If you ask "What the hell is wrong with the projector?" and go up to the control room, you find it's empty.

You are the projectionist, and you should have been up there all the time."

Colin Wilson

Despite recent advances in decoding neuronal information we are far from being able to reliably read someone else's thoughts. Recent studies of single cells in the human medial temporal lobe (MTL) suggest the possibility of offline decoding of 60% of concepts earlier presented to subjects (Quiroga et al., 2007). We here demonstrate the ability of epilepsy patients to voluntarily control the activity of a specific neuron previously found to represent a particular concept. By focusing their thoughts on a unique concept, patients were able to control a computer to make an image corresponding to that concept appear on a computer screen.

We use this experiment both to show the ability of patients to get access to areas in the brain that typically aren't easily and directly accessed, but more so, to show a unique setup by which various brain areas directly compete with each other on a sole percept. The patient is aware of the two percepts that compete and can alter the winner by somehow weighing in towards one or the other. To this date, this is the only direct reflection of competition known to us where the race towards the single attention or percept is evident and the patient gets a clear understanding of the nature of the competition between two percepts, while being able to manipulate that competition voluntarily. This, based on recordings from single units in the brain, is the ultimate competition for percept experiment one can generate shedding light directly on the topic we are studying here. This chapter, thus, holds the answers to the questions the previous chapters raised. How is competition reflected in the human brain?

Introduction

Daily life continuously confronts us with an exuberance of external, sensory stimuli together with an endless stream of internal, cognitive deliberations, plans and ruminations. The brain is challenged to select one or more of these for processing by the conscious mind. How this competition for preferential access is resolved across multiple sensory and cognitive cortical regions is not known; nor it is clear how internal thoughts regulate this competition. Using the opportunity to record from single neurons in patients implanted with intracranial electrodes for clinical reasons, we demonstrate that humans can learn to access neurons in the medial temporal lobe (MTL) to alter the outcome of competition between external images and their internal representation. Subjects looked at a hybrid superposition of two images representing explicit concepts (familiar individuals, landmarks, objects, or animals). Subjects were instructed to enhance one image at the expense of the other, competing one, while the spiking activity from four MTL units in different brain regions and hemispheres was feedback in near real-time to control the content of the hybrid. Each of these MTL units represents a single such concept in an abstract and invariant manner (Quiroga et al., 2005). Subjects rapidly learned to regulate the firing rate of group of neurons deep inside their own brain, increasing the rate of some while simultaneously decreasing the spiking rate of others. They do so by focusing their thoughts onto one image which gradually becomes more visible on a computer screen in front of their eyes. Based on the firing of these MTL units, it was possible to reliably project subject's thought to an external display.

One way of manipulating one's own thoughts is through imagination. Studies of imagination in humans have shown that neuronal correlates of the imagery process can be identified using fMRI (O'Craven & Kanwisher, 2000) or single-cell recordings (Kreiman, Koch, & Fried, 2000b). The studies also suggest a similarity in the pattern of activity to those observed when people view a visual presentation of a concept, as well as to that observed when a person views various presentation of the same concept. Prior studies have suggested that areas in the MTL encode concepts in an invariant manner (Quiroga et al., 2005), allowing for evoking of similar firing rate patterns when subjects view concepts from a similar category (Kreiman, Koch, & Fried, 2000a), or when those concepts are imagined (Kreiman et al., 2000b). Further studies exploring the degree of classification of these neurons showed that it is possible to correctly decode approximately 40 percent of the 100 selective stimuli presented to the subject in a passive viewing session using 2 to 10 of these neurons (Quiroga et al., 2007). The methods used, which were simple and linear, have shown that decoding performance is best between 300 to 1000 ms after stimulus presentation, and is typically better for neurons that are selective for a single concept.

Neurons in the human MTL are assumed to be involved in the processing of memories (Eichenbaum, Yonelinas, & Ranganath, 2007), navigational cues (Ekstrom et al., 2003), and attention (Posner & Petersen, 1990). The MTL receives substantial inputs from various sensory areas (Van Essen et al., 1992). Information processing by individual neurons in the MTL relies mainly on feed-forward pathways from the visual streams (dorsal or ventral) and elicits a response by the concept/navigation cells. To this effect, we suggested a study which

allowed for the projection of subjects' thoughts of a selected range of concepts from the large number of concepts available.

Furthermore, attention has been described as a way to allocate resources to one of several competing processes in the brain, allowing for the rise of one clear, harmonious, and coherent single thought (see split-brain patients' attention studies for further discussion of multiple processes competing in a single brain (Gazzaniga, 1968)). We here investigated a unique situation where neurons from different areas of the brain directly compete for the projection of the concept they represent.

Methods

Twelve patients with pharmacologically intractable epilepsy implanted with intracranial electrodes to localize the seizure focus for possible surgical resection (Fried, MacDonald, & Wilson, 1997) participated in this experiment. Subjects were instructed to play a game in which they had to control the display of a hybrid picture that consisted of two superimposed images using a variety of cognitive strategies (Figure 116). Four images were chosen based on a prior screening session during which we identified units that were each selective for a different image (Quiroga et al., 2005). Prior to each experiment we repeated the screening session with only the 4 selected images, to assure that the units are still responsive to the given image (“mini-screening”). At the end of each experimental block we again repeated the mini-screening in order to test for various effects of plasticity in the units. Each trial started with a 2s display of one of the 4 potential target images. Subjects next saw an overlaid hybrid image pair consisting of the target and one of the 3 remaining images ('distractor') and were told to enhance the target image (“fade in”) while diminishing the distractor image (“fade out”) by focusing their thoughts on one but not the other. The initial visibility of both was 50% and was adjusted every 100ms by feeding the firing rates of four MTL neurons into a real-time decoder (Quiroga et al., 2007) that could change the visibility ratios until either the target was fully visible (success), the distractor was fully visible (failure), or until 10s had passed (timeout; see Figures 103, 104, 105 for examples). In each block, subjects performed 32 trials, 8 for each of the 4 images, in random order. Half of the trials in each block were sham trials in which the visibility of the images did not reflect the neuronal activity during the current trial but activity from a previous trial with a different image pair as target and distractor.

See Figure 110, Figure 111 and Figure 113 for an example of multiple trials for one patient. In each experiment patients performed 32 trials, 8 for each of the 4 concepts, in random order.

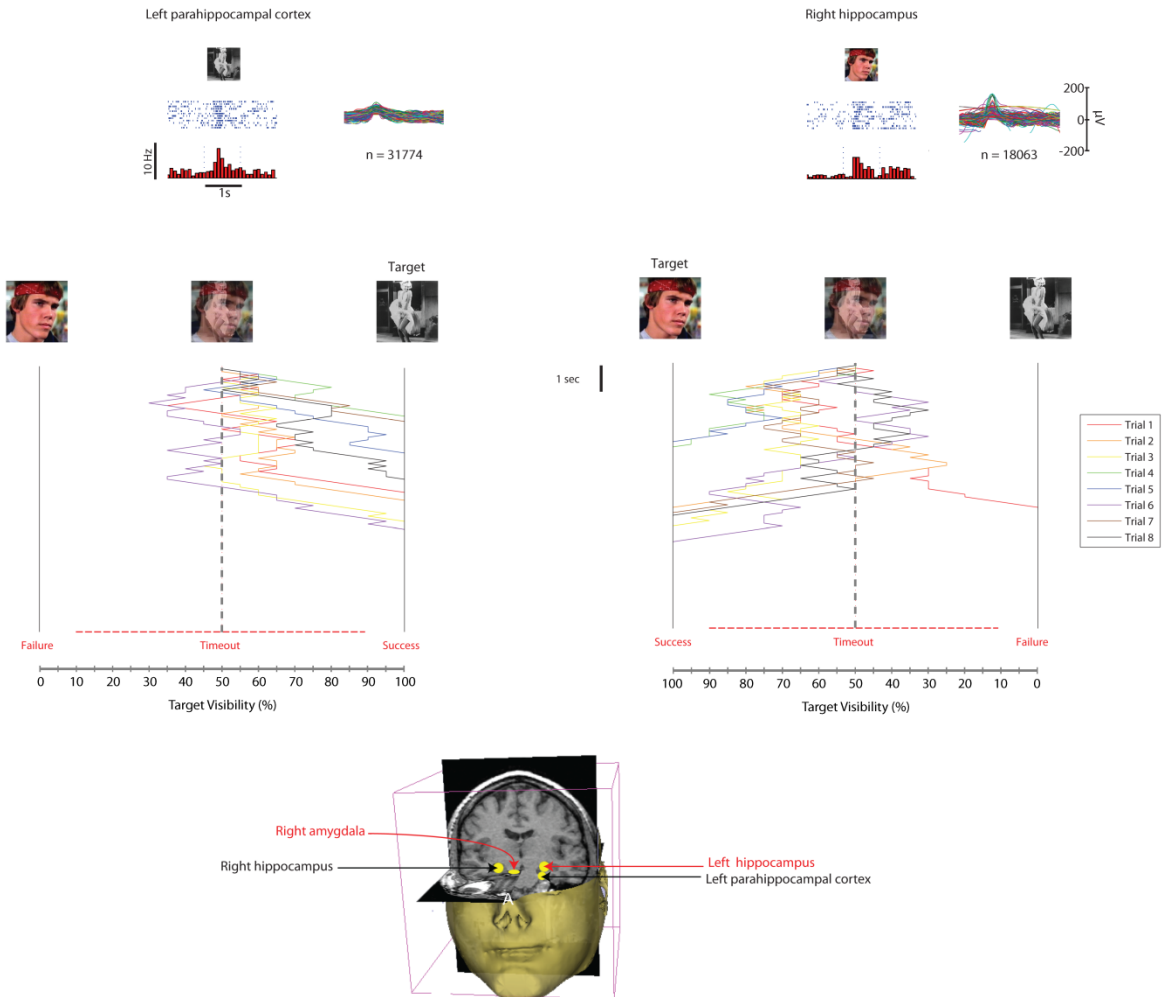


Figure 110 – Example of a single session from a patient

Out of 16 trials, an image of Josh Brolin (actor, known to the patient from her favorite movie “The Goonies”) was the target for 8 trials, and Marilyn Monroe was the target for the remaining 8. Left panel shows trials where the patient was asked to fade a 50/50 hybrid of the

two images into Marilyn Monroe. Each trial corresponds to a color-coded line of steps that, in turn, correspond to decoding of 100 ms firing rates of 4 neurons depending on which of 2 concepts was thought of by the patient (or neither). The patient was able to fade the image into Marilyn Monroe 100% of the times (although in the early part of trials 1 and 8 she initially moved a bit towards Josh Brolin and gradually figured out how to better control the signal. In the final 4 trials the patient was increasingly faster in getting to the target and could almost entirely avoid flickering between the two concepts. The right panel shows the 8 trials in which the patient was asked to fade towards Josh Brolin. Patient succeeded in 7 out of 8 trials. Similar trends of controlled fading of concepts are shown in all patients.

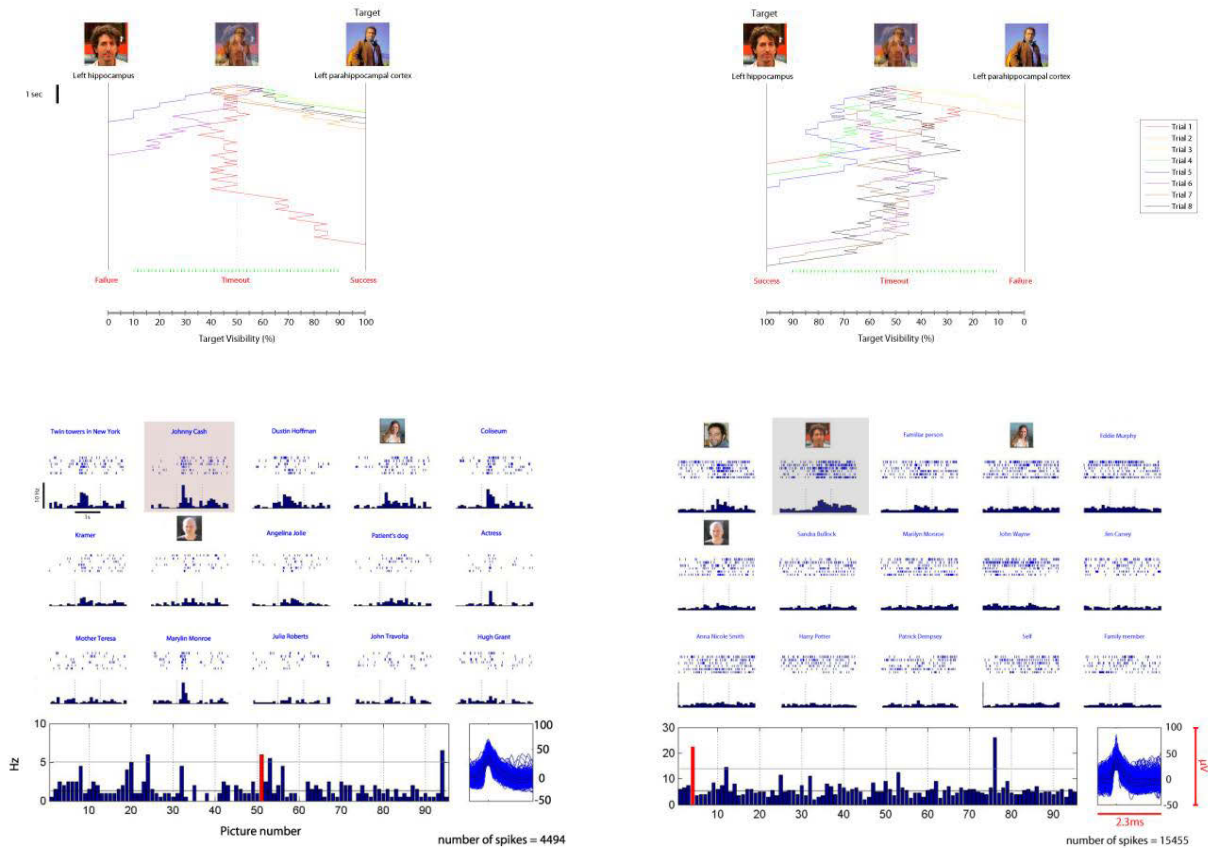


Figure 111 – Example of a single session from patient 1

Out of 16 trials, a person with whom the patient was familiar (with a corresponding single neuron response in the Left hippocampus) was the target for 8 trials, and Johnny Cash (with a corresponding single neuron response in the Left parahippocampal gyrus) was the target for the remaining 8.

Right panel shows trials where the patient was asked to fade a 50/50 hybrid of the two images into the familiar person. Each trial corresponds to a color-coded line of steps that, in turn, correspond to decoding of 100 ms firing rates of 4 neurons depending on which of 2 concepts

was thought of by the patient (or neither). The patient was able to fade the image into the familiar person 6 out of 8 times.

The **left panel** shows the 8 trials in which the patient was asked to fade towards Johnny Cash. Patient succeeded in 6 out of 8 trials.

Lower panels are the responses from the mini-screening prior to the fading experiment.

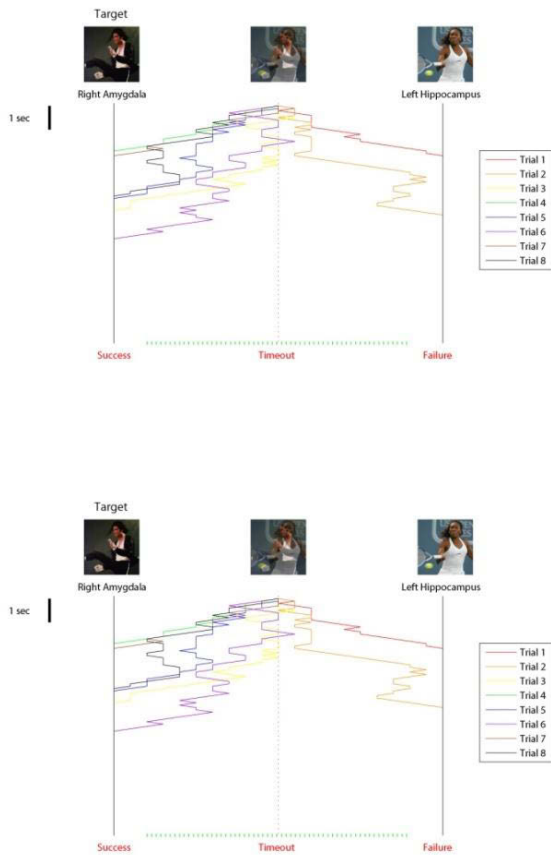


Figure 112 – Example of a single session

Example of 16 trials for patient 6, where the targets were Michael Jackson (concept corresponding neuron located in the Right amygdala) and the tennis player Venus Williams

(neuron in the Left hippocampus). Patient succeeded in 6 out of 8 trials for Michael Jackson and 7 out of 8 trials for Venus Williams.

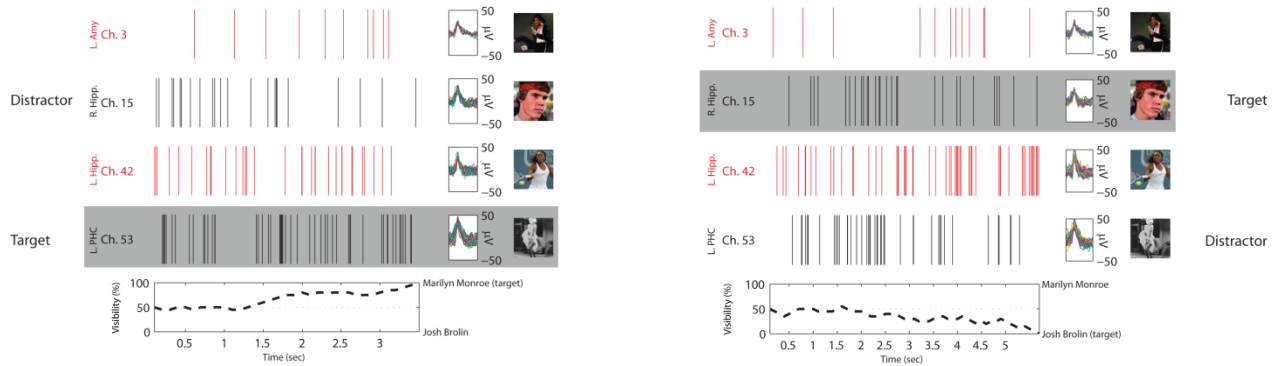


Figure 113 – Example of a single trial

Trial 1 of patient 6. Each line shows the spikes (single line) for each channel in each brain region. On the right the spikes shapes, and the corresponding concept the patient was to think of. Bottom line shows the fading “walk”, starting with a 50/50 hybrid of Josh Brolin and Marilyn Monroe and ending with full Marylin Monroe image (successful trial, as this was the target). Channels 15 (right hippocampus) and 53 (left parahippocampal gyrus), in black, correspond to the images shown to the patient in the particular trial, and channels 3 (left amygdala) and 42 (left hippocampus), in red, correspond to the channels that were not shown (decoding of those yields a “stay” in the same location).

Experimental paradigm

Screening

An initial morning screening session was recorded during which approximately 100 images of famous persons, landmark buildings, animals, and objects were presented 6 times in random order for 1 second each. A standard set of such images was complemented by images chosen after an interview with the patient which determined what celebrities, landmarks, animals and objects the patient might be most familiar with. This approximately 30-minute-long experiment (100 images x 6 repetitions x (1+R.T. seconds)) is then evaluated offline in order to tell which of the 100 images elicited a response in at least one of 64 recorded channels in the brain (based on the criteria suggested by (Quiroga et al., 2005), generally measuring the median firing rate during the 300 ms to 1000 ms after image onset across 6 repetitions and comparing it to the baseline activity of the channel. Stimuli with median firing rate 5 standard deviations above baseline are considered selective (Fried et al., 1997).

From the group of selective responses we choose 4, based on the activity. The general guidelines were: a) Choosing responses from different brain regions so as to allow for competition between regions, b) selecting responses that have similar characteristics in terms of latency and duration of the response within the 1 second the selective image is onscreen and c) choosing responses where the difference between firing rate during presentation and baseline is particularly clear. This selection is done by eye, and is not quantitative.

Mini screening

The afternoon paradigm began with a short mini-screening presentation of the 4 images in random order, 12 repetitions each, in a manner replicating exactly the setup of the earlier screening session. This allowed for a subsequent comparison of the neuronal firing pattern during the period of exposure to the stimuli, as well as for comparison of firing patterns before and after the experiment.

The data from the mini screening procedure allowed for the setup of a population-vector based decoder which was used for the decoding of the patient's thoughts in response to any of the 4 stimuli in the remainder of the experiment.

Fading

The fading experiment consisted of 32 trials – 8 for each of the 4 stimuli, which were shown in random order. Each trial began with a 2-second presentation of the target image for that trial. The patient then viewed an overlay of the target image and one of the remaining three images (these 2 images were paired for the entire first block). The hybrid images were constructed by making the 2 images transparent such that the initial state shows exactly 50% opacity of the first image and 50% opacity of the second image. The patient was instructed to “fade” the hybrid image on the screen into the target image by “continuously thinking of the concept in that image”. The patient was not directed in any way to imagine that particular image, or to focus on an aspect of the image, but was rather allowed to explore the vast area of thoughts which might elicit a response. See [Figure 114](#) for an illustration of the experimental structure.

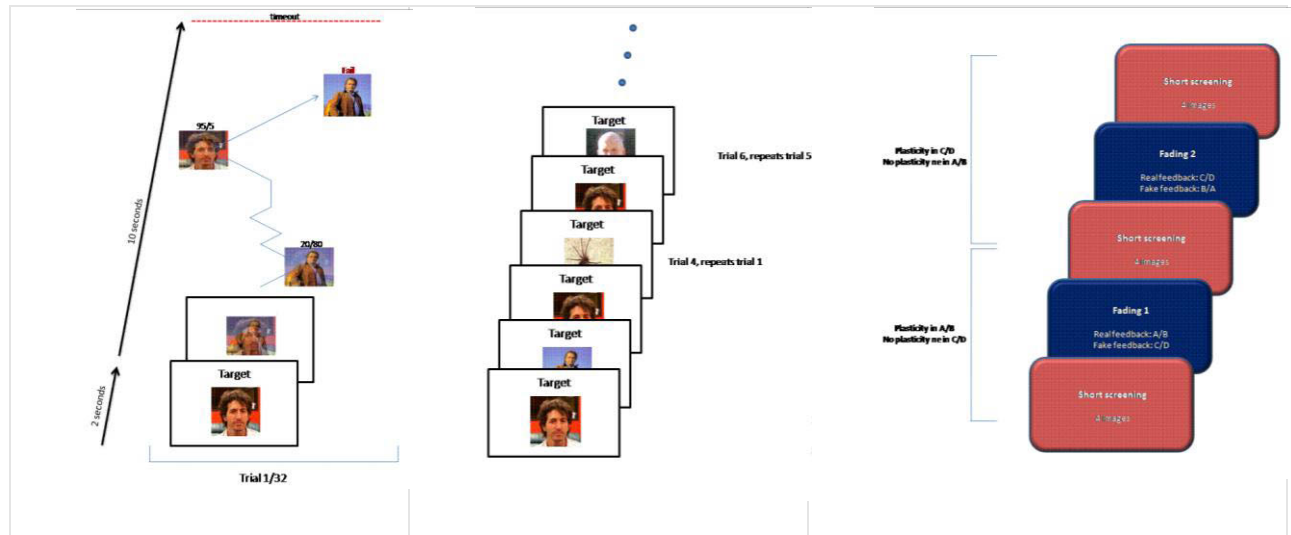


Figure 114 – Illustration of the experiment

Right panel shows an illustration of the entire experiment, broken into 5 blocks. Experiment had 3 repetitions of the mini-screenings (blocks 1, 3, 5), and 2 fading blocks (2 and 4). Real feedback in block 2 was given to two out of the four neurons, while the remaining two received fake feedback. The pairs alternated in block 4.

Central panel shows an illustration of the 6 targets in a fading block, corresponding to fading 1 in the right panel. While in the example given the patient is receiving feedback coming directly in real-time from his neuronal activity for the picture of Johnny Cash and the author, he receives false feedback (not directly coming from his neuronal activity) for the picture of the spider and the short-haired person.

Left panel illustrates a single trial in the experiment (corresponding to a single trial in the central panel), where the patient had the author as his target, after which he faded in and out of

images of the author and Johnny Cash until he reached a 100% visual presentation of Johnny Cash (“failed” trial).

At the end of the trial, acoustic feedback was given to the patient indicating a success, failure, or timeout.

Since we wanted to tell if indeed the success is due to the feedback and not the training, fake trials were used in the course of the experiment. This also allowed us to see if the potential change in the neuronal activity is due to the feedback. In the course of the 32 trials, 16 trials would have feedback for concepts A (8 trials where A is the target) or B, while 16 trials would have concepts C or D. In each fading block 2 out of the 4 concepts (say, A and B) in fact received fake feedback (that is, feedback that is not directed by the output of their immediate neuronal activity). For balanced exposure the fake trials were a direct repetition of one of prior real trials. Subsequently when the performance of the 4 neurons is analyzed, we hypothesize that the 2 neurons which received real feedback during a single experimental block would show a change in their behavior, while the 2 whose feedback is false would not show any difference.

The patient was unaware of the existence of fake trials during the course of the experiment.

Full procedure

At the end of the fading trial we repeated the mini-screening (12 repetitions of each of the four images; 1 second presentation each), in order to test differences in the behavior of the 4 neurons before the fading and after the fading based on real feedback for 2 of the 4 neurons.

After the screening we repeated the fading experiment with 32 trials that seem the same as the

previous fading to the patient, but in fact accurate feedback was given for images C/D, and false feedback was given for images A/B. At the end of the second fading block, we repeated the mini-screening for the 3rd time to validate the responsiveness of each neuron for a particular concept throughout the experiment.

Decoding

Data from four selected channels, i.e., microwires, are read, and spikes are detected in real-time for every 100ms interval during the mini-screening. Each 1s image presentation in the mini-screening (4 images x 12 repetitions) is broken into ten 100ms bins. We use only the 7 bins from 300ms to 1000ms for the analysis as these include the most relevant data for decoding (Quiroga et al., 2007). The total numbers of spikes for each 100ms bin form clusters in a 4-D space representing the activity of the four units for each image. Thus, for 12 (repetitions) x 4 (images) x 7 (bins) we obtain a 336 (cluster) by 4 (channels) matrix corresponding to the firing rate during each image presentations for all 100ms bins (Figure 115).

During the fading, the firing rates from the four channels gave rise to a population vector that was used to associate the corresponding 100ms bin to one of the four images. The population vector was regarded as a point in 4-D space, and we used the Mahalanobis distance to determine which cluster the point was closest to: Mahalanobis distances were chosen as distance measure since this is a fast and linear distance calculation measure that takes into account the shape of the cluster (previous data showed that clusters variability is significant for our data (Quiroga et al., 2007), thus taking the standard deviation of the cluster into account yielded better decoding). The equation for the distance calculation is:

- $x = (x_1, x_2, x_3, x_4)$ is the new point in the 4-D space (corresponding to the firing rate of 4 units in the previous 100ms).
- S is a 336×4 matrix of firing rates of 4 units during 100ms bins in the mini-screening when subject was viewing one of four images (e.g. columns 1:7 in the matrix corresponds to seven 100ms bins of the firing rates of the 4 channels while image A was on the screen, columns 8:14 correspond to activity while image C was on the screen, etc). \bar{S} is the mean of S .
- The distance from each of the 4 clusters is thus calculated by:

$$D = (x - \bar{S}) \times \text{COV}(S)^{-1} \times (x - \bar{S})^T$$

- $D = (d_1, d_2, d_3, d_4)$ where d_i corresponds to the distance from cluster i . the point x is classified with the cluster with the lowest distance.

The closest cluster is regarded as the one the sample belongs to, and as the concept the subject is thinking of. Notice that each trial consists of two concepts that, when decoded, directly influence the visibility of the two associated images that make up the hybrid (annotated as A and B). Decoding of one of the other two concepts (annotated C and D) is interpreted as ‘thinking of neither A nor B’. In any given 100ms of each fading trial, there are 3 possible outcomes: (i) The sample is close to the cluster representing image A, causing the opacity of image B to increase by 5% and the visibility of A to increase by 5% in the hybrid image seen by the subject. That is, if the proportion of transparency of images A/B was 50%/50% in the previous 100ms, it will now change to 55% of A and 45% of B. (ii) The sample looks more like a sample in the cluster associated with image B, which would lead to a 5% fading in the

direction of image B. (iii) The outcome is that the sample looks more like images in clusters C or D. This does not result in any change in the hybrid image.

Any one trial can last as little as 1s (10 consecutive steps from 50%/50% to 100%/0% or 0%/100%). A limit of 10s was set for each trial, after which the trial is regarded as ‘timeout’ even if the subject was closer to the target or the distractor.

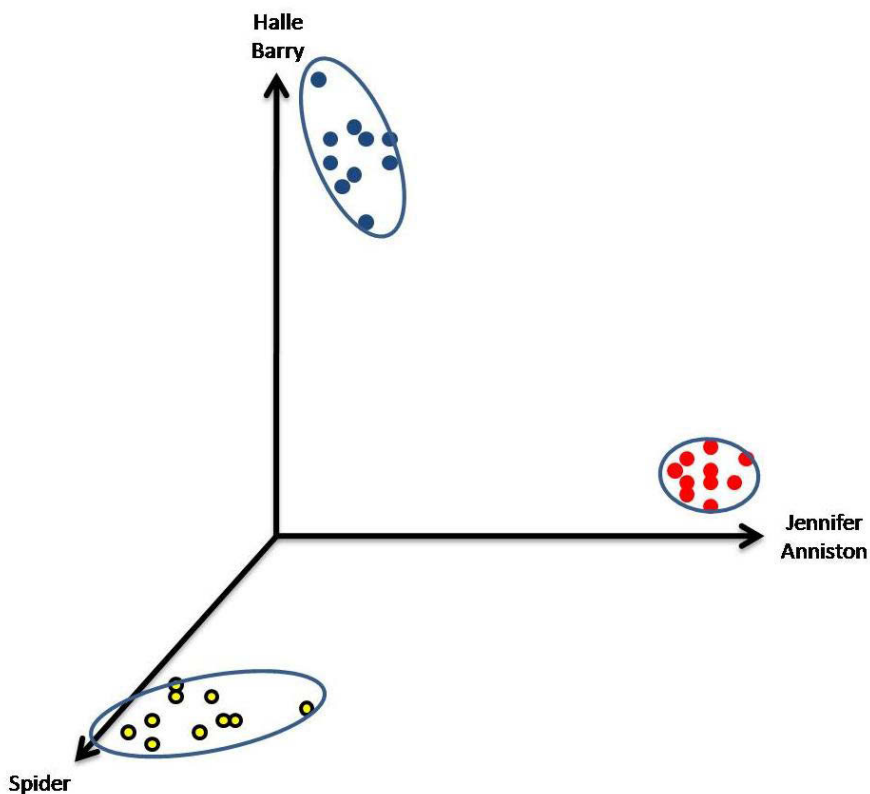


Figure 115 - Illustration of the decoder

For 3 neurons selective for 3 concepts (Halle Barry, Jennifer Anniston, and a spider) we would look at the data as coming from orthogonal 3D space where each axis is the firing rate of the corresponding neuron. During the mini-screening we would fix one point for the firing rate of

the 3 neurons during the presentation of a picture of Jennifer Anniston, same for Halle Barry, and for a spider. Since during the mini-screening each picture is shown 12 times we would gather 12 points for each concept. More so, each 100 ms out of 7 (300 to 400, 400 to 500, ... 900 to 1000) will be given a separate point to allow for even higher resolution of identification of the thought at any given 100 ms bin. The clusters rising from the showing of the 3 (in our experiment 4) concepts will then be fixed and from then on each new population vector will be classified as belonging to each of the clusters and accordingly the thought decoding will be done.

Setup

The experiment is run on a 15' laptop computer with 160 x 160 pixel-sized images centered on the screen at a distance of about 50 cm from the patient. Data from the patient's brain is acquired using the Cheetah system (Neuralynx, Boise, MN), from which it is sent via TCP/IP to a server performing the spike detection. On the spikes detection server the 4 selected neurons are filtered (band-pass at 300–3000 Hz), and a threshold is applied to detect spikes. This threshold is set prior to the experiment based on a 2 minute recording of each channel while the patient is sitting still with eyes opened. Spike counts in the 4 channels, per 100 ms bin, are then transferred via TCP/IP to the experiment laptop computer where the data is used for the online control of the faded images. The experiment was programmed using Matlab (Mathworks, Palo Alto, CA) and the Psychophysics toolbox, while the spikes detection proprietary software is written in C++ for efficiency and real-time analysis (code provided on authors website at <http://www.klab.caltech.edu/~moran/feedback>). See Figure 116 for illustration of the experimental setup.

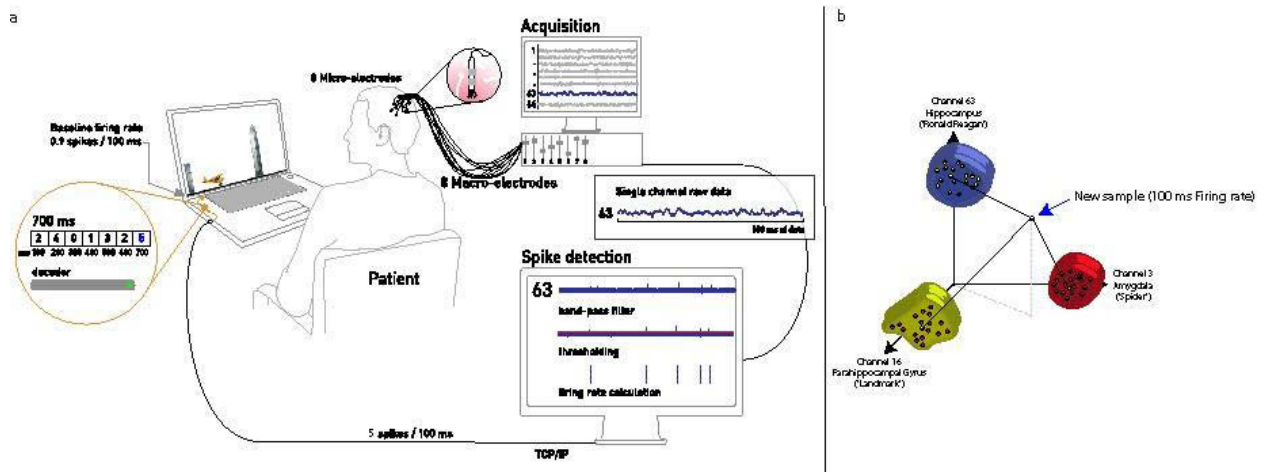


Figure 116 – Illustration of the experimental setup

a. Data recorded from 8 different macro-electrodes, each having 8 microelectrodes and 1 reference, in the medial temporal lobe in the patient's brain is transferred via 64 channels to the acquisition system. Four pre-selected channels from 4 responsive units are transmitted to a server performing real-time spike detection and firing rate calculation. A 4-D vector, corresponding to the number of action potentials in the previous 100ms from those 4 units, is sent to another computer. The decoding algorithm running on this machine determines the composition of the hybrid image presented to the subject. The total delay between spike detection and determination of the visibility of the hybrid is < 100ms.

b. Every 100ms, the 4-D vector is tested to determine its closest Mahalanobis distance (distance weighted by the standard deviation) to one of the 4 clusters representing one of the 4 images. If the 'winning' cluster represents the target or the distractor image, the visibility ratio of these two will be adjusted accordingly

Sites

We analyzed units from the hippocampus, amygdala, entorhinal cortex and parahippocampal cortex.

We recorded from 64 microwires in each session, acquiring data from up to 6 single neurons per microwire. We identified a total of 133 units (68% multi-units and 32% single-units) that were responsive to at least one picture. Out of these responses we selected 4 in each of the 18 sessions. Out of these responsive units, 58 multi-units and 14 single-units were used in the subsequent fading experiment (see Figure 117 for a distribution of the units used).

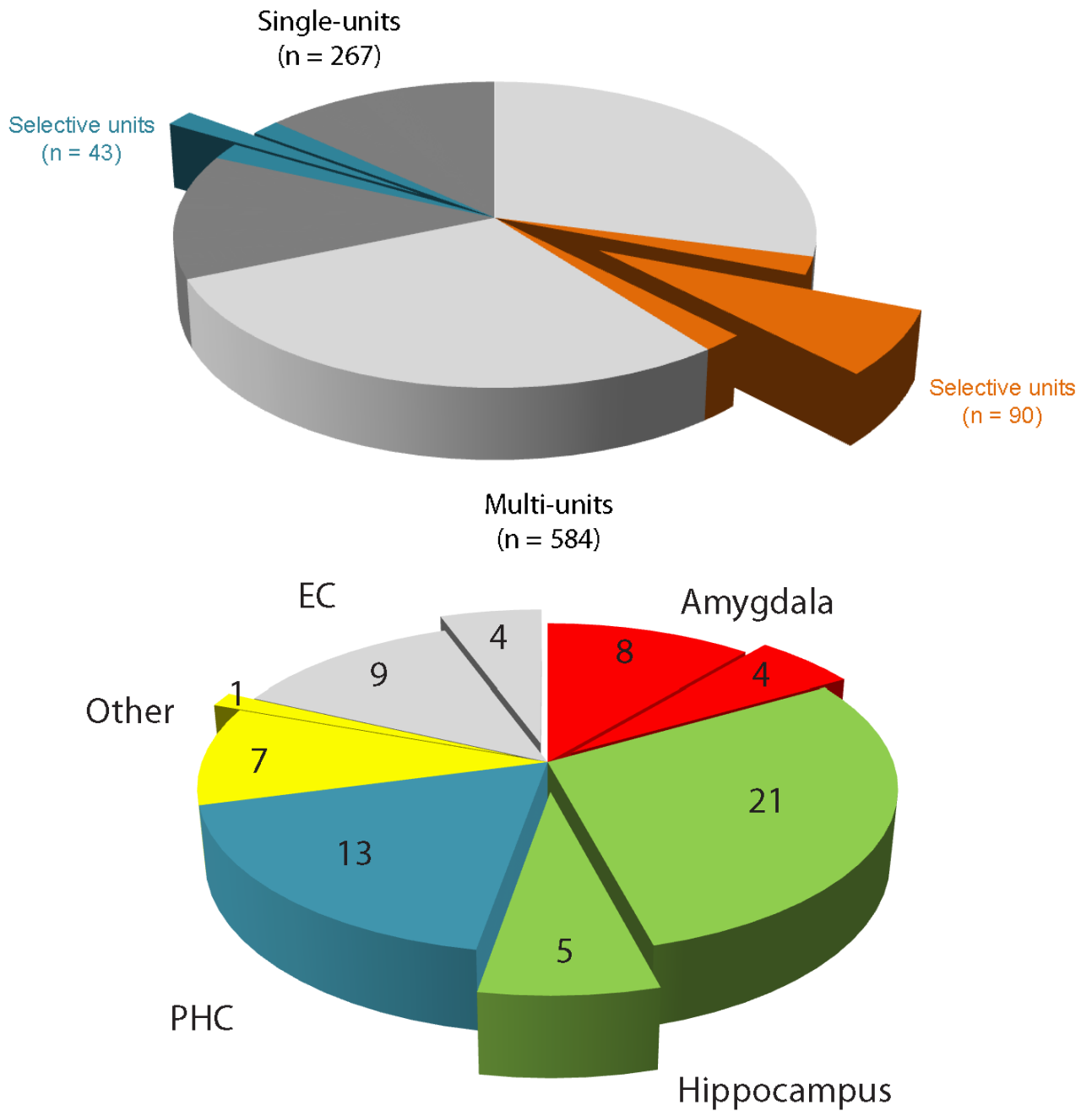


Figure 117 - Units distribution

- a. Out of 851 units recorded in the course of 18 sessions, with 12 subjects, we identified 584 multi-units (69%) and 267 single-units (31%). Out of these, 133 (90 multi-units and 43

single-units) were responsive to one or more images in a prior screening. From these selective units, we used 58 multi-units and 14 single-units in the fading trials.

- b. The 72 units used in the fading experiments, distributed by regions. The exploded slices represent single-units for each region. Annotation: PHC - parahippocampal cortex. EC - entorhinal cortex.

Responses were either positive (exhibiting an increase in the firing rate above baseline), or negative (decreasing the firing rate). Excitation was determined using the following techniques developed in previous works (Quiroga et al., 2005), by considering a wider interval after picture onset (100ms to 1000ms) and adding a paired t-test using $\alpha = 0.05$ as significance level. Inhibition was determined using the following four criteria: i) the median number of spikes in the interval (100ms to 1000ms) after picture onset was at least 2SD below the baseline activity, ii) a paired t-test using $\alpha = 0.05$ as significance level rejected the null hypothesis of equal means, iii) the median number of spikes during baseline was at least two, iv) the median difference between the number of spikes in the response interval and the baseline interval was higher than the background activity of 95 randomly resampled responses (bootstrapping). Only responsive cells were kept for further study. See table 1 for a breakdown of the responses.

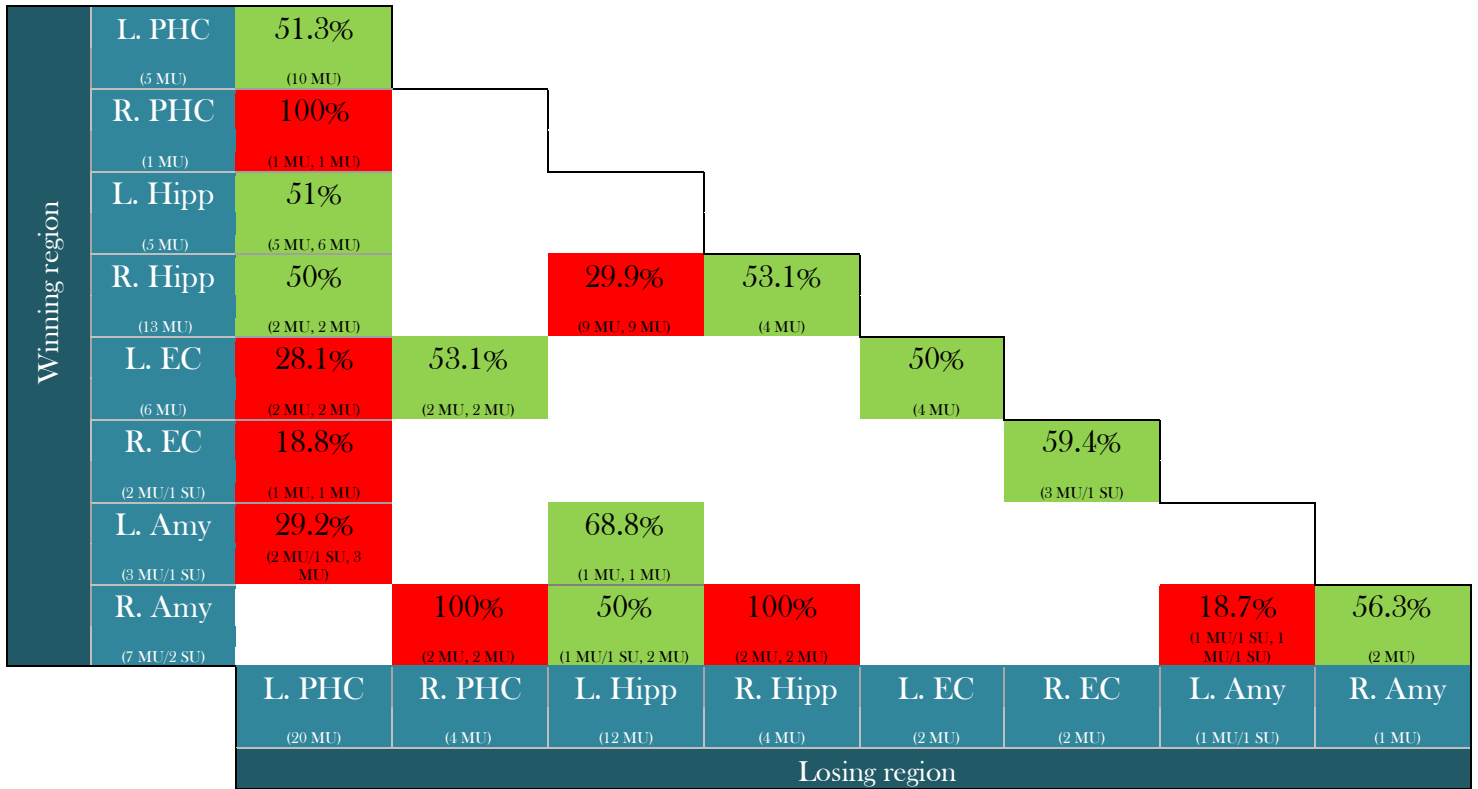


Table 9 – Competition across brain regions

A breakdown of the percentages of wins for each brain regions pair. Each pair totals in 100% (e.g. left hippocampus wins over left parahippocampal cortex 51.0% of the times, and left parahippocampal cortex wins over left hippocampus in the remaining 49.0%). In brackets are the amount of units used in each competition (i.e., the 50% of wins of right amygdala over left hippocampus reflects overall competitions between 1 multi-unit and 1 single-unit in the amygdala winning over 2 multi-units in the hippocampus). The total number of units for each region is shown on the heading lines underneath the region name. Cells highlighted in red reflect significant ($p < 0.05$) wins or losses of the region over the competing region, using a sign-test.

Annotation: L/R - Left/Right. PHC - parahippocampal cortex. Hipp - hippocampus. Amy - amygdala. EC - entorhinal cortex.

Multi-units versus single-units

We used single-unit for most of the analysis based on neurons although the experiment was based on multi-units, as these provide better understanding of the neuronal level and reflect the actual activity by which the brain and the patient operate. Results which correspond to the actual success/fail rates, or that are based on the performance and activity which had occurred in the patient room are analyzed based on the multi-unit, as these were the actual numbers that were used to create the trials.

The purpose of the fake trials

While the fake trials in our analysis are used to counter the effects of inhibition and as a control for various questions asked in the experiment, these were not introduced mainly for this reason. The original purpose of the fake trials was to isolate the effect of plasticity on the results. We expected to see changes in the neurons that were indeed receiving real feedback in the course of a block, while we expected not to see one in neurons that received fake feedback. This would alter in the following block where the 2 pairs of neurons receiving real/fake feedback would alternate. However, multiple analysis trying to identify effects of plasticity (by 5 measures: latency, duration of response, peak of response, time to peak, and average number of spikes) showed no repeated effect in our patients – thus making the original purpose of the fake trials obsolete. While we feel that a behavioral change should manifest itself in some sense on the neuronal level, the resulting hypothesis from our lack of shown plasticity is that the differences are either reflected in a different area, thus not showing plasticity in our neurons, or – more likely – that the changes are diminished when we test them, as our tests

are always in the 'passive' screening configuration. As such, even if a neuron does alter its behavior in the fading block, it may as well show the original stereotypical activity when again tested under the screening conditions. However, none of these claims can be in fact answered using our current data.

Bootstrap testing of statistical significance for task performance

In order to compare the performance of individual subjects (as in Figure 119) to a chance level, we set individual baselines in the following way: each subjects' sequence of 32 trials (8 trials x 2 real images x 2 fading blocks) was broken into its individual 100ms steps, such that the decoding result for each step was categorized as 'to target', 'to distractor', or 'stay'. For example, in the first trial (colored red) on the left panel of figure 2a (where the target is Marilyn Monroe) the first six 100ms steps are 'to target', the seventh 100ms step is 'to distractor', the eighth is 'stay', and so on. Each subject, thus, ended up having a total number of bins reflecting the proportions of steps he or she used during the course of the entire experiment. This proportion reflects the subject's own baseline chance of going in either direction (the subject in Figure 103, for instance, had altogether 389 steps where she went towards the target, 49 steps towards the distractor, and 18 'stay' steps). Using these proportions we generated 1000 new 32 trials blocks. For each 100ms step, we randomly generated a direction of movement based on the probabilities calculated for each subject, and then generated trials. For each block we calculated the performance and then compared the 1000 realizations to the one the subject's actual performance. If the subject's performance were based only on his/her personal biases (moving in a certain direction because of faster response onset by one unit, paying more attention repeatedly to one of the two competing concepts, etc.) then the random realizations

should exhibit a similar performance. The subject's actual performance would be better than the random realizations only if the subject was able to use his or her moves accurately to maneuver the fading of the two images towards the target.

Image saliency

When images are first presented to subjects in the beginning of a trial, each image's visibility is 50%. However, some images might be more dominant than others even when balanced in such a way because of their coloring or content. This could affect the initial direction of a trial, as the dominant image might draw more attention from the subject, causing an initial movement towards it. While it is hard to tell which image draws the attention of our subjects more objectively, we tested all images pairs viewed by our subjects with our standard saliency model (L. Itti & C. Koch, 2001). We computed a saliency map for the first 3 fixations on the 50%/50% hybrid image. This allowed us to verify that no pair had in one of the 3 most salient locations a patch of any of the images that is more visible than the other (e.g. for the combination of Marilyn Monroe and Josh Brolin images shown in Fig 103, when computing the standard saliency map model, no patch either from the Brolin image or from the Monroe image is drawing attention in the first 3 fixations). This suggests that none of the images we used was significantly more attractive than its counterpart.

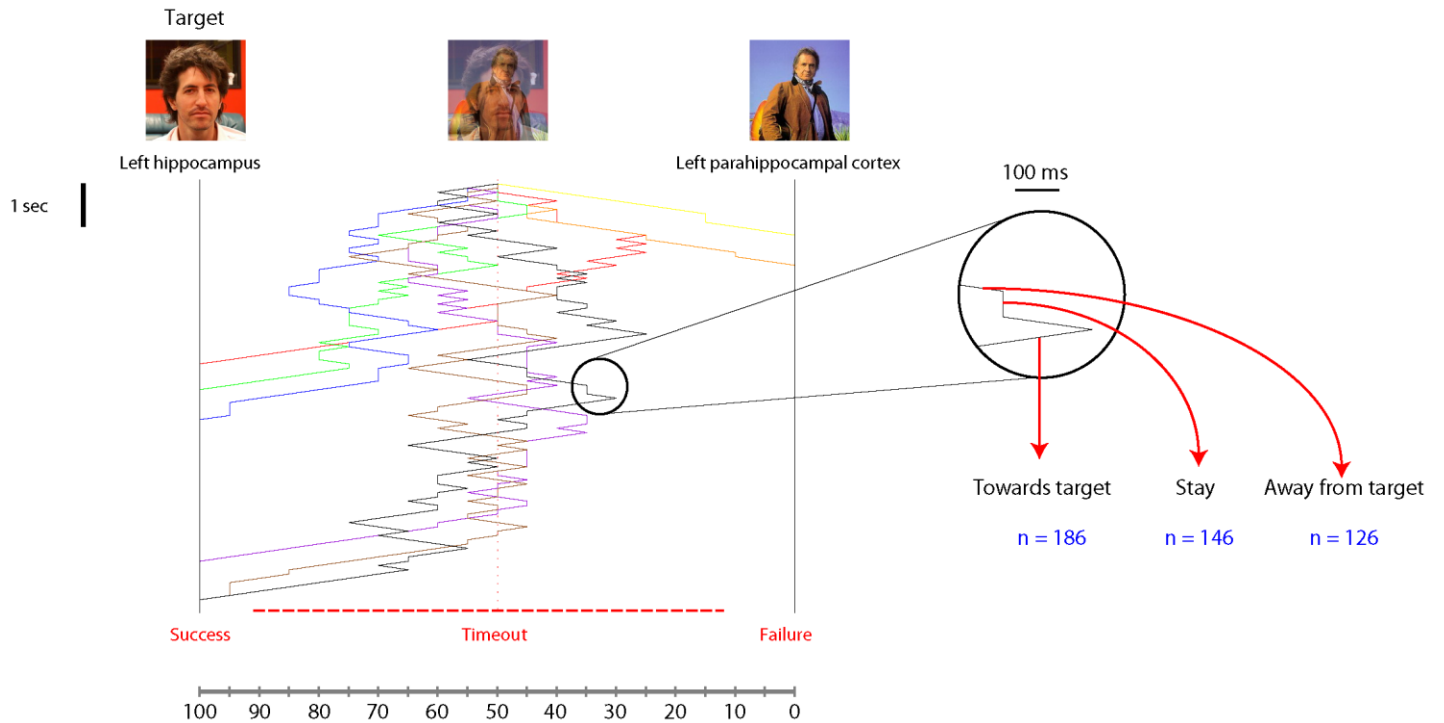


Figure 118 - Illustration of the bootstrapping technique

Left panel shows the fading path for 8 successful trials in another subject. Each 100ms step was classified as one of three possibilities: 'towards-target', 'away from target', or 'stay'. Notice that 'towards-target' steps can occur even in cases where the subject was seeing mainly the distractor on the screen. The subject had 186 steps towards the target, 126 towards the distractor, and 146 stays. With this proportion he was able to successfully reach the target in the majority of cases (6/8). For the bootstrapping we used these proportions to create 8 new trials and compared the performance in those to our results. We repeated this procedure 1000 times and tested how many of the 1000 trials indeed showed lower performance than ours (typically 3-4 successful trials out of 8 for these particular proportions). Notice that in the given case out of 186 'towards-target' 100ms bins, the average firing rate was 3.02SD above baseline (0.86 Hz), while the 'stay' and 'towards-distractor' were less than 0.89SD above baseline. Notice that

the subject was successful in the majority of cases (6/8) even though the distractor composed the majority of the hybrid image. The fading was governed by the subjects' voluntary manipulation of his neurons' firing rates rather than the visual input (i.e., the subjects could think of the target, while the distractor was mostly viewed on the screen, distinguishing the sensory system from voluntary thought).

Results

Subjects could manipulate the visibility of the hybrid image by any cognitive strategy of their choosing. Six out of 12 subjects reported in a follow-up interview that they primarily relied on imagining the concept represented by the target picture (most often a person) or closely allied associations. Other subjects used a variety of methods, including suppressing the distracting concept. We looked at the “trajectories” of movement in the continuous plane defined by the image from the 50%/50% hybrid towards either the target or the distractor, and analyzed the trial lengths, the time spent viewing one dominant image on the screen while thinking of its counterpart and actively altering the nature of the competition within a subject’s brain, and the ability of subjects to control the activity by activating single neurons.

Subjects participated in the task without any prior training and with a striking success rate in a single session that typically lasted around 30 minutes. We recorded from a total of 851 units, out of which 72 responsive units were used for feedback (in a post-hoc analysis detailed in the supplementary materials we analyzed the units used in the experiment, and identified 58 multi and 14 single-units. However, during the experiment we did not perform any spike clustering in order to maintain the real-time decoding speed under 100ms). Overall, subjects succeeded in “reaching” the target image in 596 out of 864 trials (69.0%; 23.4% failures and 7.6% timeouts). To test the performance of each individual separately, we used a bootstrapping technique - generating trials by randomly shuffling 100ms segments of activity for each of the four units and testing performance for these chance trials. Results were significant ($p < 0.001$) for each of the twelve subjects (Figure 119). When the first two trials from each subject's block were assumed to be de facto training trials, and we analyzed the results from the remaining six,

performance improved by 3.8% to 72.8%. Subjects succeeded in manipulating the hybrid image to the target in their first trial in 124 out of 208 first trials (59.61%).

It came as a surprise to us that subjects could control which one of two images dominated within a single session, that is, within a few minutes. As success involves a combination of suppressing the distractor image and enhancing the target, we monitored the degree of learning in two complementary ways, focusing on whether subjects took longer to fail and/or became faster at winning. Figure 4a plots the time-to-fail for one subject who, in a competition pitting a picture of the cyclist 'Lance Armstrong' against a picture of one of the lab members, failed in all 8 trials. Strikingly, the time-to-fail lengthened over the 8 trials (see Figure 104 for another example). Testing for all subjects showed that for all 8 blocks where subjects failed in all trials of a single target the time-to-fail increased as subjects gained more experience (Figure 122; $p < 10^{-5}$, Spearman's rank correlation). The average slope of the trial times fitted to a linear model is 0.89 ± 0.21 s/trial. That is, after each failed trial, subjects took 0.89s longer to fail on the next trial. Finally, time-to-fail as compared to the previous trial increased in 99% of successive failed trials (and not just in those with 8 consecutive failed trials). In the 12 blocks where subjects successfully reached 100% visibility of the target image in all 8 consecutive trials, no timing difference is apparent (Figure 122). The difference between consecutive successful trials was -0.14 ± 0.67 s (n.s; $p = 0.42$, Spearman's rank correlation). This might be due to a floor effect, i.e., the time-to-success on the first trial is already close to the theoretical minimum of 1s (as the visibility is updated by 5% every 100ms, starting from 50%). We conclude that successful manipulation of neural firing can be achieved rapidly, and often within the first trial. In those cases where subjects failed to control neuronal firing, they at least learned to delay failure.

To understand the neuronal mechanisms driving the competition between the stimuli we tested the changes in firing rate for the 4 units used during the trials. We calculated firing rates in 100ms bins (the window length for each decoding step), and assigned these to one of 3 categories: approaching the target (moving from, e.g., 65%/35% target/distractor visibility to 70%/30%), approaching the distractor (e.g., moving from 65%/35% to 60%/40%), or no change (Figure 118). Approaching the target could be the result of inhibition of the distractor unit, or of excitation of the target unit. We calculated the baseline firing rate of each unit on the basis of the firing rate 700ms before each image presentation in the prior mini-screening session (Quiroga et al., 2005), and compared the average firing rate during the average 100ms bins to that of the baseline in order to detect changes in the average firing. In 84.6% of successful trials, a combination of excitation (3.56SD above baseline; $p < 10^{-4}$ for every subject, Wilcoxon signed rank-sum) of the target unit and inhibition (2.84SD below baseline, $p < 10^{-4}$, Wilcoxon signed rank-sum) of the distractor unit led to success. In the remaining successful trials, the effect (inhibition/excitation alone) was not significant. No deviation in firing rate was seen in the two other units we recorded from during each block. That is, success *in the fading experiment was not due to a generalized excitation or inhibition of all neurons.*

To disentangle the effect of visual feedback from that of the mental state of the subject, we contrasted the activity of units at the same visibility levels across different trials types (Figure 124). We compared the activity of each unit in successful trials when the target was the unit's preferred stimulus (target trials) with activity in successful trials when the target was the unit's non-preferred stimulus (distractor trials). This comparison was always done for the same level of visibility, that is, for the same visual feedback. We normalized each unit's response by its

maximal firing rate over the entire block, and averaged all trials for all subjects and all blocks. We see a significant task-related difference in the average firing rate for each of the 19 visibility ratios (5%, 10%, ... 95%) ($p < 10^{-4}$, Wilcoxon signed rank-sum with Bonferroni correction). The unit firing rate is higher when subjects focus their thoughts on the target, even for the same visual stimulus. That is, in a trial where the visual feedback from the hybrid is 80% of image A and 20% of image of B, and the subject was told to enhance A, the unit responsive to A fires above baseline. In a later trial, the visual feedback is the same, but this time the subject was asked to focus on B, the A selective unit will fire less than baseline. The only difference was the mental state of the subject, following the instruction to suppress one or the other image. This was not the case for failed trials (mean p value is 0.18).

We decoded the target on a trial by trial basis purely by the firing rate of a unit. That is, for 92% of successful trials, we could identify the target unit purely by testing whether the average firing rate of one out of four units we recorded increased above its baseline in all 8 trials for a given image. This clear change in neuronal firing rate based on the internal mental state of the subject indicates that the neuronal feedback was given by the subject's thought and not by the external stimulus.

Excitation of the target unit, alongside inhibition of the distractor unit, occurs even in trials where the distractor is dominating the hybrid picture, suggesting that the units are driven by voluntary thought capable of overriding distracting sensory input. To test for the mechanisms that govern this, we directly compared vision and imagery. Out of the 235 (27.19%) trials where at some stage of the trial the distractor had a higher visibility than the target, the subject was

able to eventually win 71.7% of those trials. That is, the composite image shifted back towards the target, despite the distractor being more visible than the target. If fading is entirely controlled by vision, these trials would be expected to end in a loss. To test the significance of these winning trials, we bootstrapped trials based on the subjects' proportions of 100ms bin that shifted toward or away from the target image (see above) and compared them to those of trials where the distractor was dominant. This demonstrates that the majority of such cases would have ended in failure, instead of actually being successful ($p < 0.01$, Wilcoxon signed rank-sum, shuffling trials based on the proportions starting with the a-priori bias towards the distractor).

To further quantify the extent to which successful game playing was due to the visual feedback from the changing hybrid image, we compared performance during normal feedback to that reached during sham feedback. Such trials are a replay of a previously recorded sequence of spikes using a different target/distractor pairing. As subjects had not been told about the sham feedback, their level of effort and attention were the same as during real feedback. Success dropped dramatically from 596 to 270 out of 864 trials (31.2%; 33.7% failures and 35.1% timeouts). This difference is highly significant ($\chi^2 = 69.9$, $df = 2$, $p < 10^{-4}$). Only 2 out of 12 subjects performed significantly above chance ($p < 0.001$); the rest were not significant (p values: 0.15 ± 0.14 , mean \pm std). Furthermore, neither improvements in performance nor the increased delay of failure observed during real feedback (Figure 122) were evident during sham feedback (mean increase in time between successive failed trials -0.09 ± -0.78 s, $p > 0.20$; mean speedup for successful trials -0.13 ± 0.63 s, $p > 0.40$). Finally, we analyzed neuronal control during sham trials by binning each 100ms step for each trial by the outcome of the decoder.

Unlike for real feedback, no consistent trend was seen in the firing rates of either target or distractor units nor a reduction or enhancement of firing rates during sham feedback. These findings support the notion that feedback from the four selective units controlling the composite image were essential to successfully carry out the task, rather than the general cognitive efforts of the subject, exposure to the stimuli, or global changes in firing activity.

Is there any long-lasting effect of feedback on the excitability of the MTL neurons? That is, do those neurons whose firing rate was up- or down-regulated by subject's thoughts retain any chronic changes in their responsiveness? We used five criteria to test for changes in neuronal activity in the mini-screenings before and after the game: latency, duration, peak firing rate, mean firing rate, and time of peak of activity of the individual neurons. No significant change was seen in any of these parameters. This suggests that either the feedback had no lasting effect on the neurons, or any sustained effect is not apparent when subjects are exposed to the images in passive viewing during our mini-screening procedure.

Our experiment creates a unique design to interrogate the brain's ability to influence the dominance of one of two stimuli by decoding the firing activity of four units deep inside the brain. The stronger the target neuron fires, the more visible the target and the more opaque the distractor in the resulting hybrid picture on the screen (and vice versa). Overall, subjects successfully fade 69% of all trials. Better spike discrimination might have increased performance, but would have slowed down feedback to unacceptable levels. Cognitive processes voluntarily initiated by the subject, such as thinking about the target image (e.g. 'Lance Armstrong') or suppressing the distracting one rapidly and simultaneously affects the

firing activity of units in different brain regions, sometimes even across hemispheres (see supplementary materials for list of all regions).

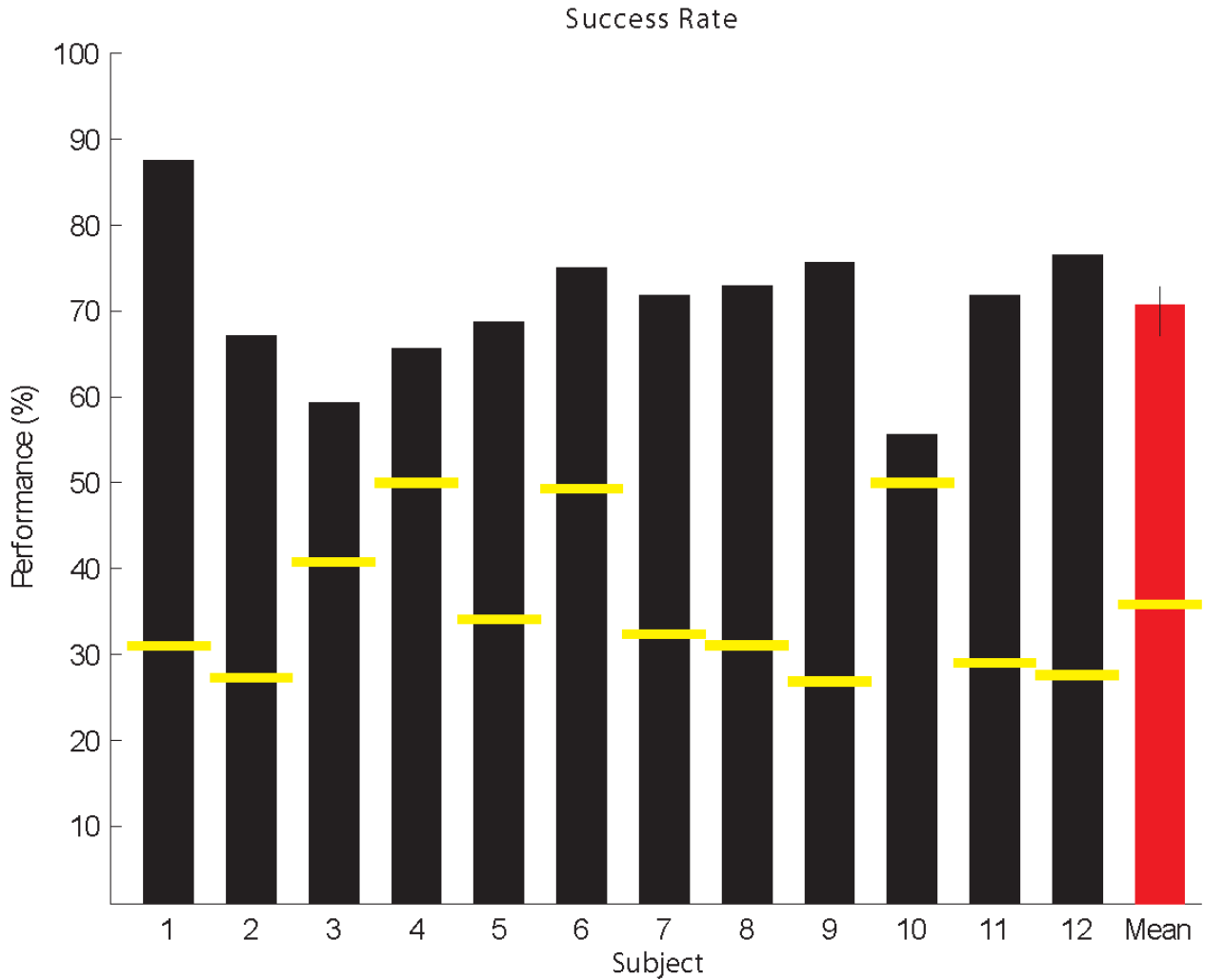


Figure 119 - Success rate

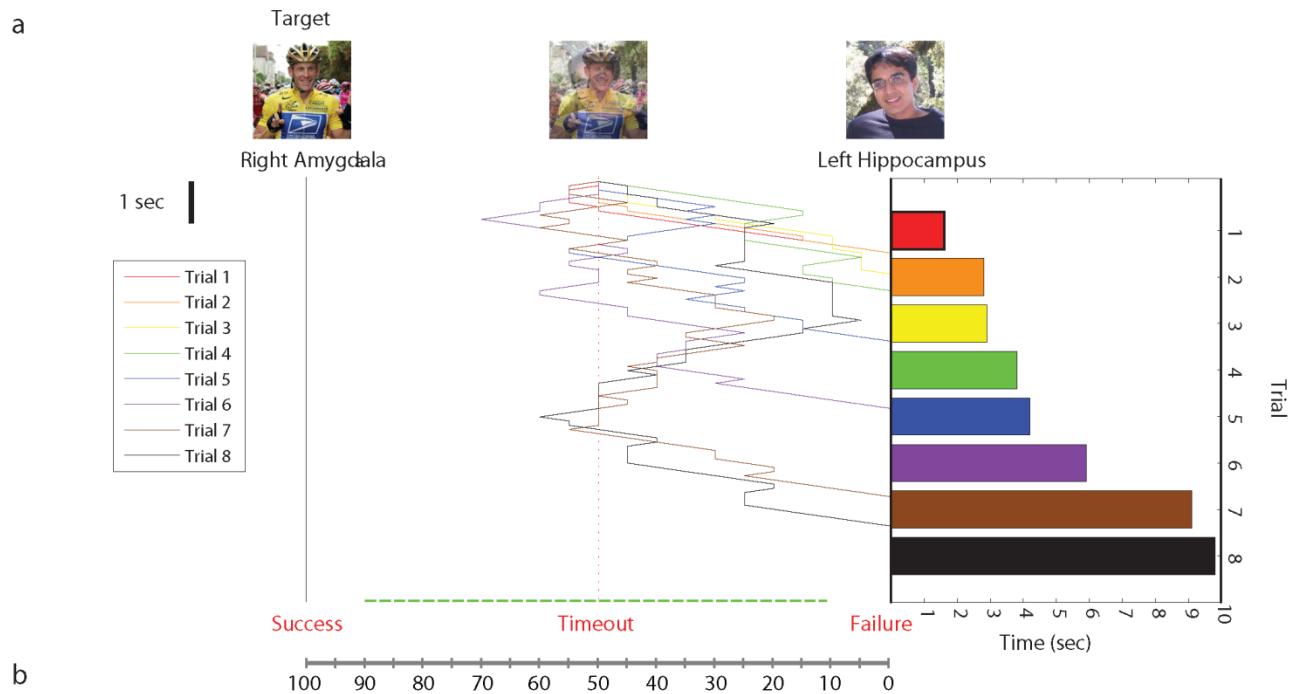
Bars represent the percentage of trials in which subjects successfully controlled the activity of 4 units and faded to the correct target. Yellow lines show chance levels - determined by generating 1000 random trials for each subject (Monte-Carlo). Right red bar is the mean of the

performance averaged over all $n = 12$ subjects, and the mean of the chance level performances.

These results are significant with $p < 0.001$.

In order to examine the patients' improvement throughout the experiment we looked only at the failed trials for a given patient and measured the amount of time it took for the failure to occur. We expected the time to failure to lengthen with practice. Our reasoning was that even when patients were unable to reach the target before timeout, they would be more successful at delaying the failure by enhancing their thoughts of the target. Failure in these cases occurs mainly due to shorter latency of the distractor that causes the trial to fade towards it before the patient is able to initiate his focus on the target concept.

Figure 120 shows the time to failure for a patient that failed all 8 trials for a given concept. As predicted, the time to failure lengthens with time.



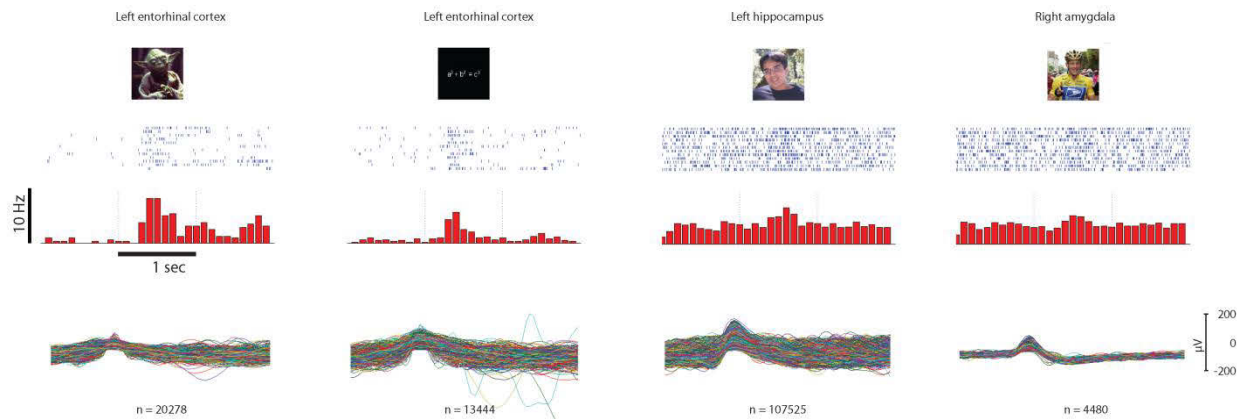


Figure 120 – Learning for a single patient

a. Example of learning in a single subject. The fading walk diagram for subject 2 in Figure 119 who failed in all 8 trials when ‘Lance Armstrong’ was competing against one of the lab member’s pictures. On the left is a corresponding color-coded bar diagram of the time-to-fail - defined as the time to the complete visibility of the distracting image. While the first trial failed in a mere 1.60s, failure was delayed for 9.80s in the last trial (timeout occurs after 10s). Color-coded bars on the right correspond to the times of each trial. See additional example in Figure 113.

b. The corresponding raster plots (twelve trials are ordered from top to bottom) and post-stimulus time histograms taken from the mini-screening prior to the block shown in panel (a). Vertical dashed lines indicate image onset and offset (1s apart). On the right are the spikes shapes throughout the entire session for each of the four units. The two units from the left entorhinal cortex (‘Yoda’ and the ‘Pythagoras theorem’) were used in the following block. Notice that the ‘Lance Armstrong’ response is quite noisy when the spikes are not sorted (i.e., when we use the multi-units rather than a single unit determined afterwards by the spike

shapes), which might explain the poor performance in this session. Nevertheless, the subject was able to alter the signal-to-noise ratios during the fading trials, thereby delaying failure more and more.

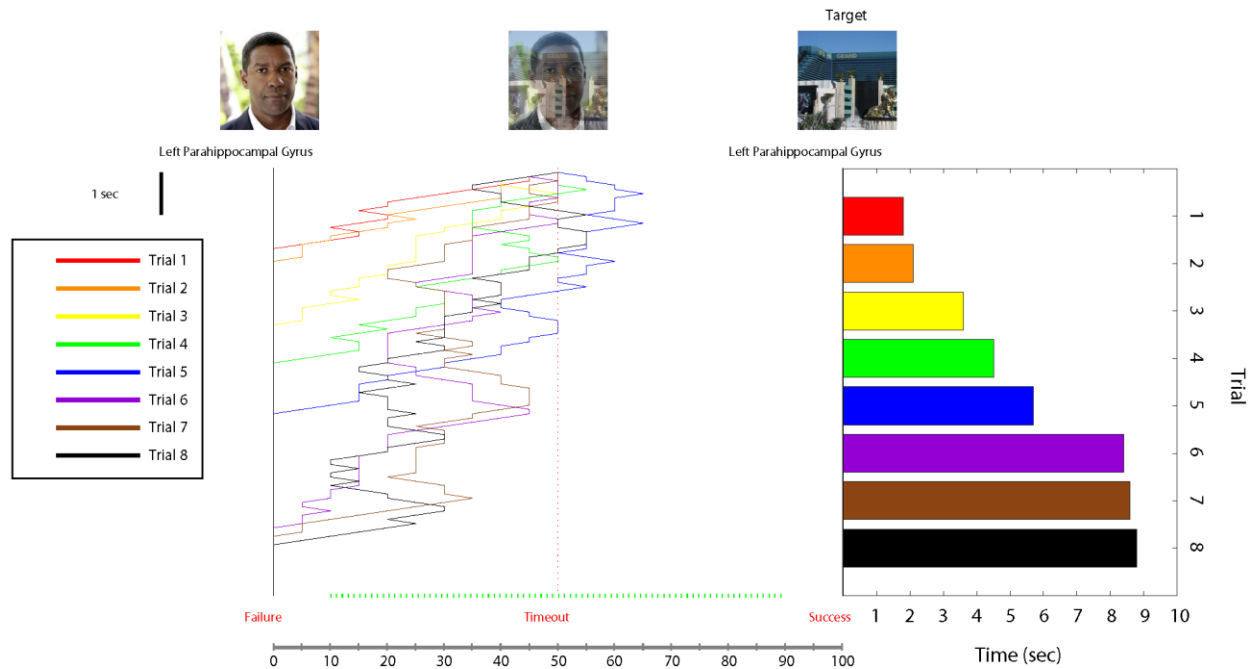


Figure 121 - Additional example of a learning effect

The subject failed to enhance the image of a building in Las Vegas that the subject was familiar with 8 times. However, as the bar graph of times on the right shows - the subject was able to continuously delay the failure.

Testing for all patients (Figure 122) shows that indeed such learning (be it of the network, neuron, or patient) increases the trial duration ($p < 10^{-2}$, Spearman's rank correlation).

Successful trials show hardly any difference in timing (mean trial time change was -0.13

seconds), suggesting that indeed the increase in time is due to a failure prevention process rather than training.

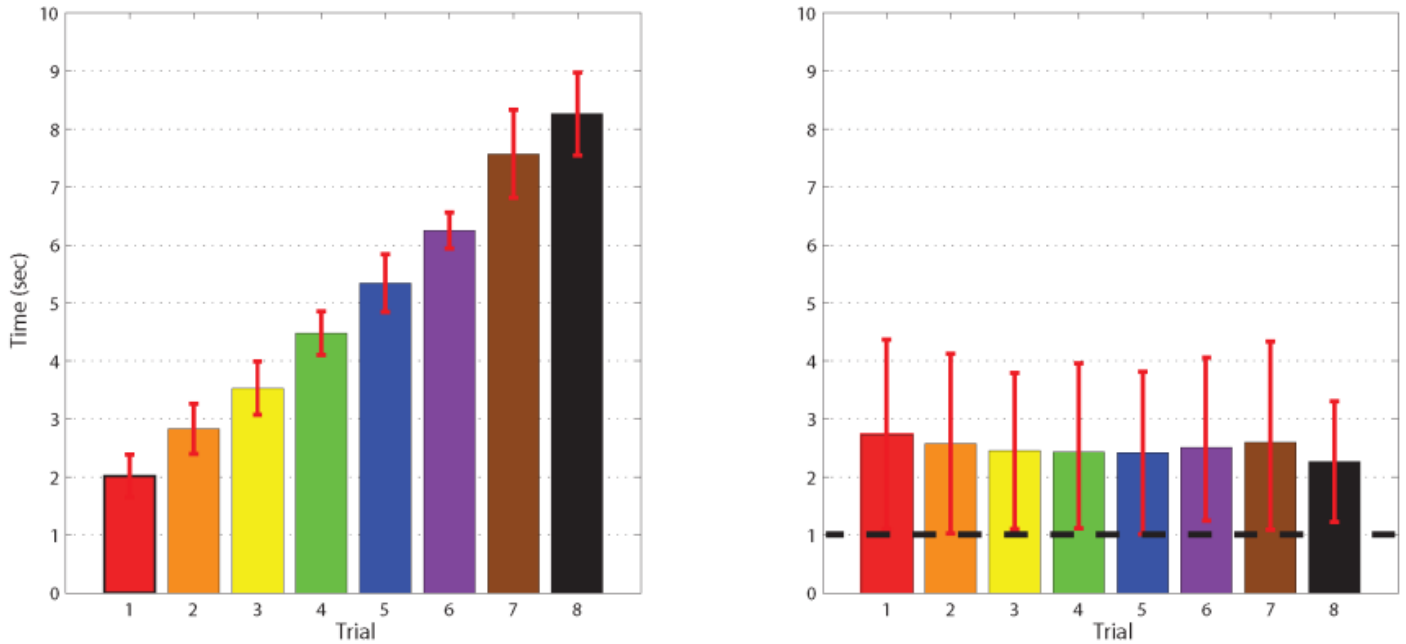


Figure 122 - Group “learning”

Left: Average of the total trial times for all 8 block in 6 subjects with 8 consecutively failed trials. X axis indicates trial number and Y axis the mean time-to-fail for each trial. Red error-bars indicate std. **Right:** Average of the total trial times for 12 blocks in 7 subjects who had 8 consecutive successful trials. While time-to-fail increases significantly, time-to-success remained constant. The thick dashed black line at 1s indicates the minimum trial length possible.

Sum of the total trials time for all the sessions that included at least one where one of four stimulus fadings failed all trials (4/10 patients). On the x axis we see the trial number (8 trials total), and on the y the length of the sum of each trial.

Next, we analyzed the neuronal control of the experiment. We were interested in understanding whether the neuron firing is correlated with the amount of the image visible on the screen or if it is purely determined by the patients' thoughts. We assigned each 100 ms of the experiment (the window length for each decoding step) to one of 3 bins: approaching the target, approaching the distractor, and no movement (which indicated a failure to decode either concept). Approaching the target could be the result of inhibition of the distracting neurons, or of enhanced activity of the neuron representing the target. We show that indeed the fading is governed by the concept-encoding neuron firing 2.2 standard deviations above baseline, which results in fading toward the target. Notice that this occurs even in trials where the distractor is dominating the hybrid picture (see Figure 123), suggesting that the neuron is driven by voluntary thought which is capable of overriding distracting sensory input. Calculating for all patients showed that for a total of 768 trials, the increase above baseline for the target PSTH is 7.7 times, while the distractor was 0.7 ($p < 10^{-4}$, Wilcoxon).

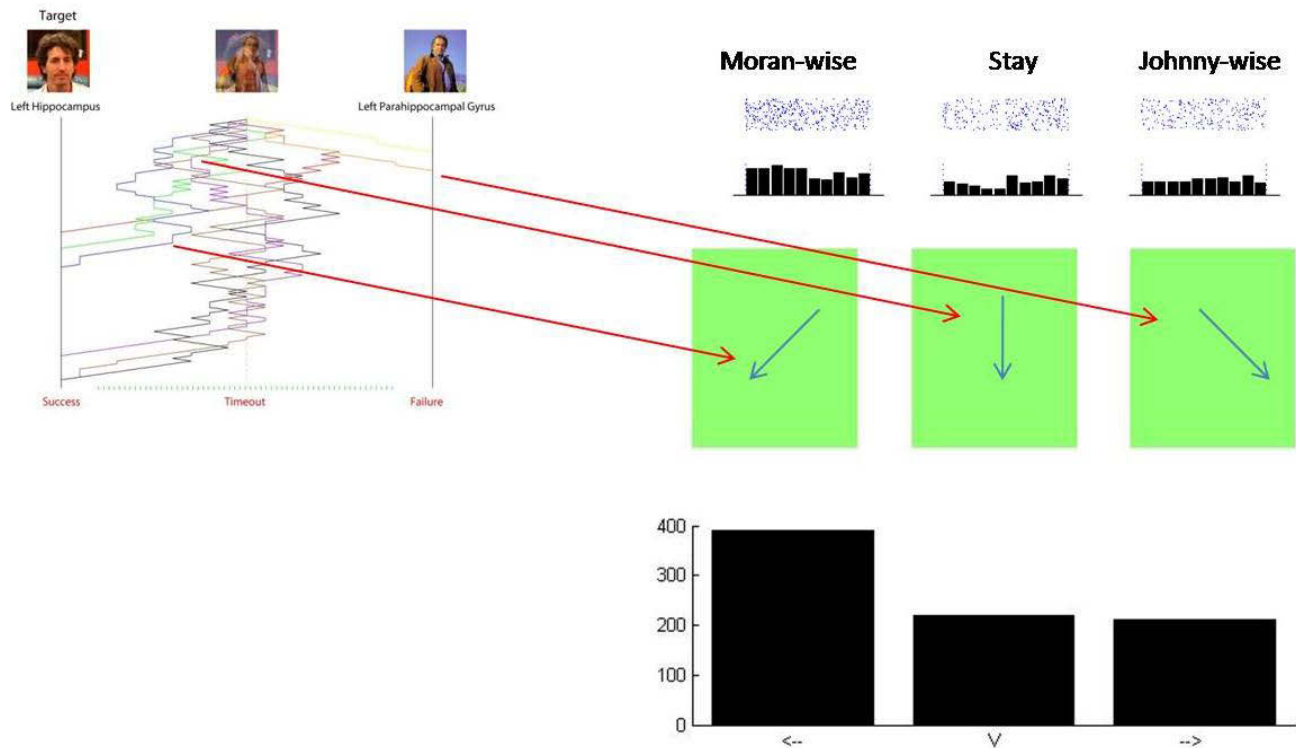


Figure 123 – Neuronal activity dependence on imagery/vision

Left panel shows the fading path for 8 successful trials. Each 100 ms step was binned into one of three possibilities: towards-target, towards-distractor, stay. Notice that towards-target steps can occur even in cases where the patient was seeing mainly the distractor on-screen.

Right panel shows a histogram of the spikes for the 3 cases. Out of thousands of towards-target 100 ms bins the average spike count was 2.1 times above baseline, while for the stay and towards-distractor these were only 0.1 and 0.08 above baseline, as shown in the sum-of-spikes diagram below, showing that the fading was governed by the patients' thoughts rather than the visual presentation (i.e., the patients could think of the target, while the distractor was mostly viewed on the screen, thus distinguishing the sensory system from the action one).

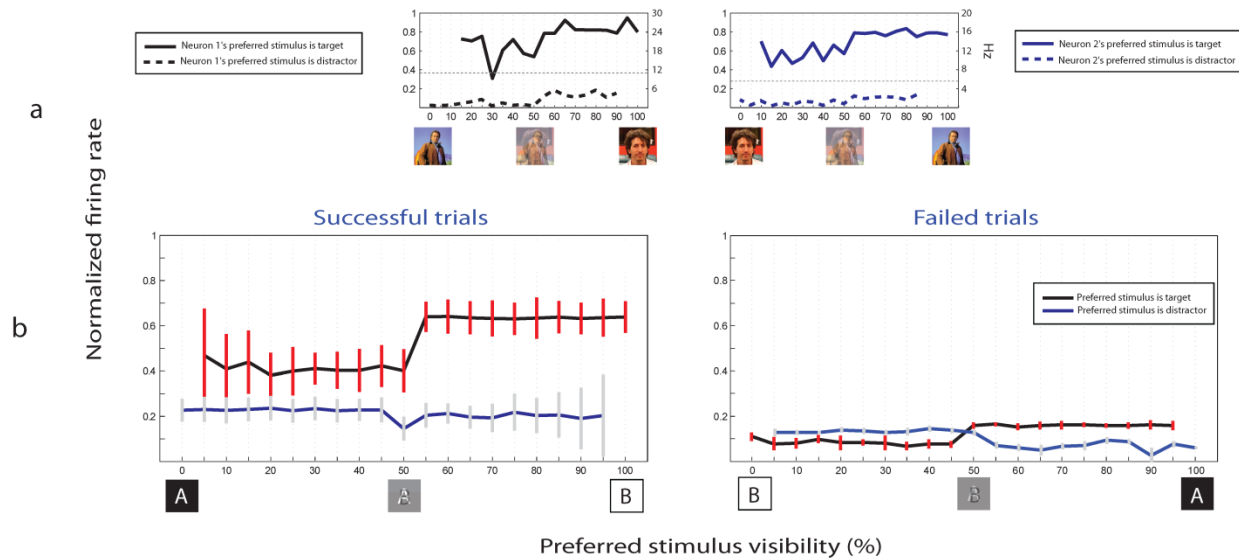


Figure 124 – Mental control of the feedback

a. Normalized firing rates of a single-unit in the left hippocampus, responsive to images of the first author, and a single-unit in the left parahippocampal cortex, responsive to Johnny Cash (see Figure 111), as a function of visibility. Taking all successful trials where the target was the unit's preferred image, we averaged the firing rates every 100ms for every level of visibility (e.g., all bins where visibility was 40%/60%). The solid lines are the firing rates in trials where the target was the preferred image for the neuron, and the dashed lines are trials where the target was the neurons non-preferred image. Both units fired significantly more (and above their baseline firing rates of 11.43 Hz for the left neuron and 5 Hz for the right neuron, marked as a grey dashed line) when the target was the preferred stimulus, and less than baseline when the target was the non-preferred stimulus. Right y-axis shows the absolute firing rates for the two neurons.

b. Averaging target and distractor trials across all subjects and all units reveals that in successful trials (left) the firing rate is significantly higher when the target is the preferred stimulus than when the target is the non-preferred stimulus, no matter the visual input. That is, when the hybrid picture seen by the subject is composed of 70% of the target image and 30% of the distractor image, the normalized firing rate of the unit is 0.63 when the cell's preferred image is the target and 0.19 when it is not. This is not the case for the failed trials (right). Red and dark grey vertical lines are standard deviations.

Invariant representation of a concept

While it is evident that in order to perform the task patient must be able to either replicate their exact configural firing rate of the 4 neurons during the presentation of each of the images on a given time, it is yet well worth in fact showing that patients do have an invariant representation of the concepts we use. Figure 125 shows a single example corresponding to the neurons in Figure 123 showing the firing rate of the neuron during the screening part of the experiment without clustering.

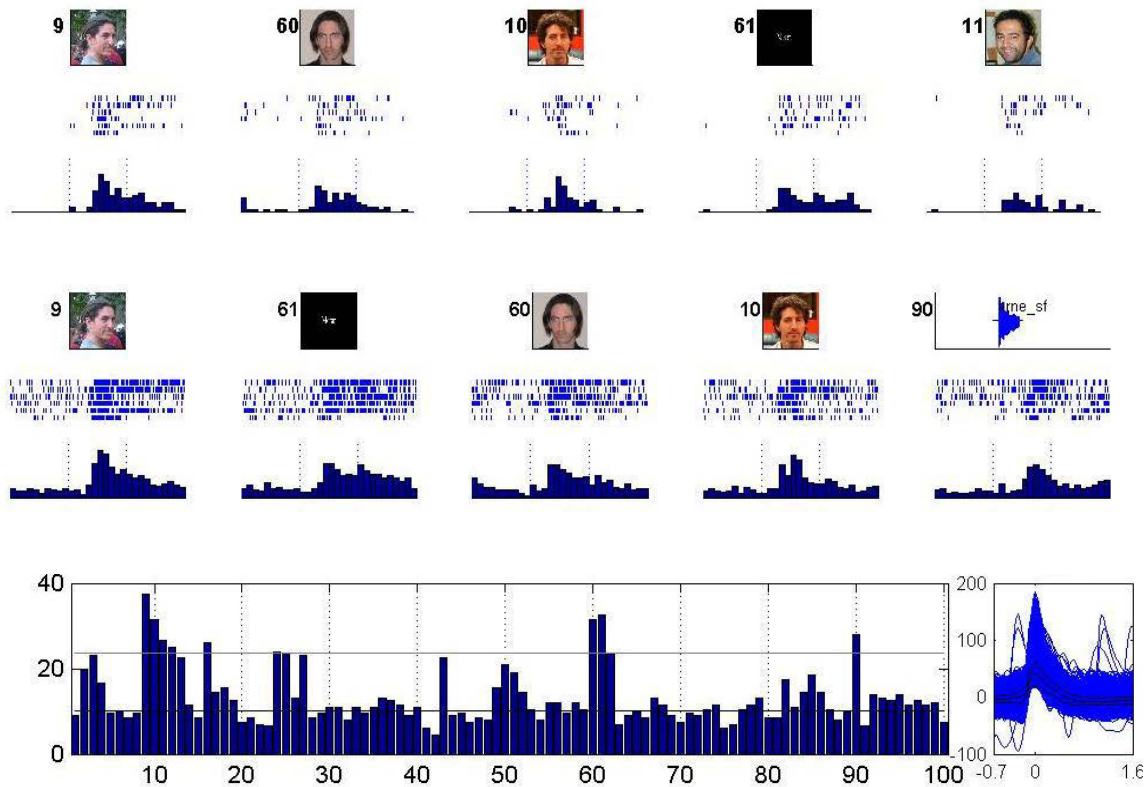


Figure 125 – Responses of channel 45, left anterior hippocampus, in patient 399

Top panel. Responses of the neuron to the five best stimuli in a screening session where multiple presentations of a single concept were shown. In this case, the five presentations (including the written name of the person) yield a selective response. These responses are based on spikes sorted manually.

Middle panel. Responses of the neuron in the same fashion as the top panel, only for non-sorted spikes (multiunit). Note that in the real time experiments, we always used the multiunits for speed, as clustering would have taken longer and would hurt the real-time nature of the experiments.

Lower panel. Left. Bar graph showing the 100 images and the firing rate they elicited. The gray line represents 5 x baseline firing rate of the neuron. Most presentations of the concept for which the neuron was selective (images 10 to 16) yield an increased activity. **Right.** The spikes plotted between -0.7 ms before spike peak, and 1.6 ms after the peak.

Comparison of real/fake feedback

In order to quantify the level of success a patient reaches at baseline we also compared the performance of each individual to that reached if the performance was based on the activity during a fake trial. While fake trials are in fact a replay of a previously generated trial, the neuronal responses are generated from the patients' effort and level of attention as they are unaware of the fact that the feedback is based on data from previous trials. Thus, their responses can be used to generate the trials that would have occurred had the feedback been based on their responses. A comparison of the performance (defined as total of failed trials subtracted from the total of successful trials) for the 32 fake trials for each patient with its 32 real trials yields a p value below 0.001 (Wilcoxon) for all patients.

LFP analysis of the data

The data analysis below corresponds only to patient 405, which had 4 'active' channels (3, 15, 42, 53) corresponding best to images (3, 1, 4, 2) or (Venus Williams, Michael Jackson, Marilyn Monroe, Josh Brolin).

Main questions addressed:

1. Is there a significant difference in the observed LFPs between the different screenings? Previous analysis indicated that the difference in mean firing rate of spikes is not significantly different across the different screening sessions.
2. Is there a significant difference in the observed LFPs between different concepts? Previous analysis by Alexander Kraskov indicated that image category can be reliably inferred from the LFPs. The question here is whether individual concepts can also be classified?

We first started by trying to quantify the variance in the LFPs in the time domain. Example of the noisy signals can be seen in Figure 126. Each subplot shows 1 sec before image onset until 1 sec after image offset. The various colors correspond to the 12 repetitions of the image. The mean signal is plotted as bold black line. This type of analysis is similar to the one by Kraskov (Kraskov, Quiroga, Reddy, Fried, & Koch, 2007).

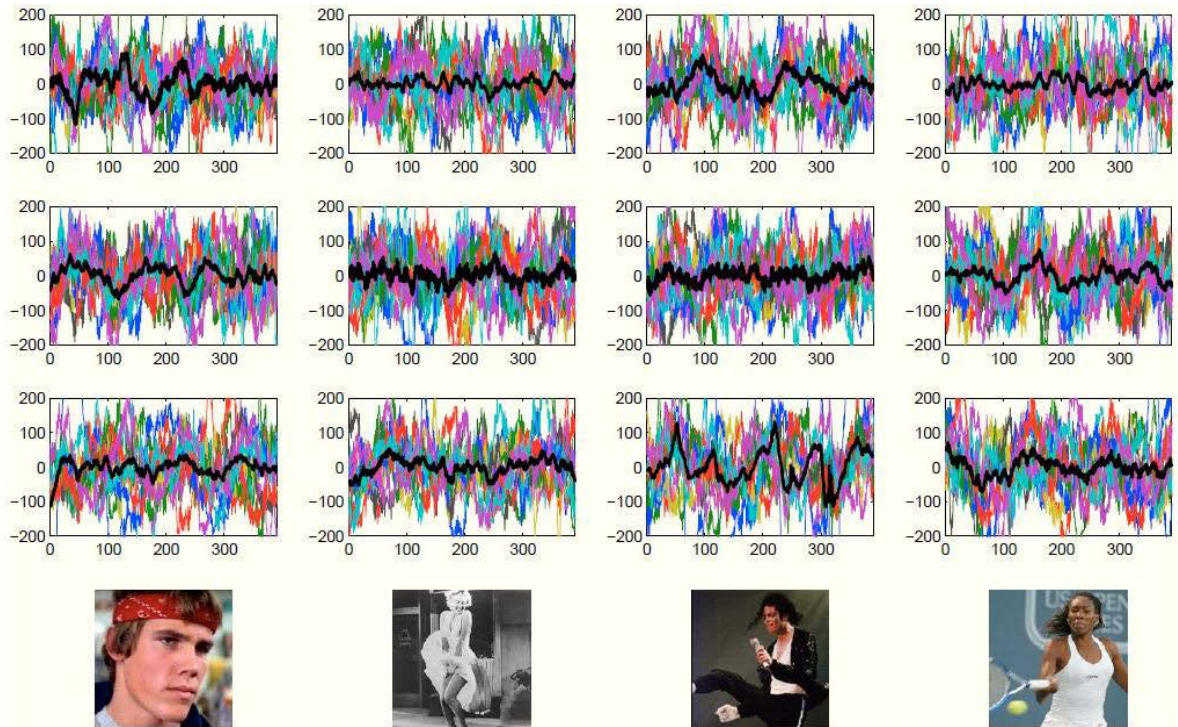


Figure 126 – LFP traces from patient 405, channel 53

Each row corresponds to a different session. Each column corresponds to a different image.

One thing to note is that the LFPs are very noisy. It seems like they are out of phase, which might explain why there is a big spread while the mean lies near zero. Also, in some cases there are outliers that need to be removed. This channel corresponds best to Goonies image (first column). A summary of all mean LFP traces, when considering all "active" channels from that experiment is shown in Figure 127. One way to continue this type of analysis is to do some point-wise significance tests and determine whether there is a significant difference across sessions. However, it is quite obvious that in some cases there is a big difference (see, Image 4, channel 42, or Image 1, channel 15), while in others this difference is much more subtle and appears only toward the end of image onset (at about 280 in the shown axis, Image 3 at channel 3, or Image 2 at channel 53). It looks to me like there is little chance of actually

obtaining something meaningful in the time-domain due to the phase differences. Analysis in the frequency domain can be done in two ways. One is to consider the short interval between image onset and image offset (or the extended one that includes 1 sec before and 1 sec after) and try to estimate the observed frequencies.

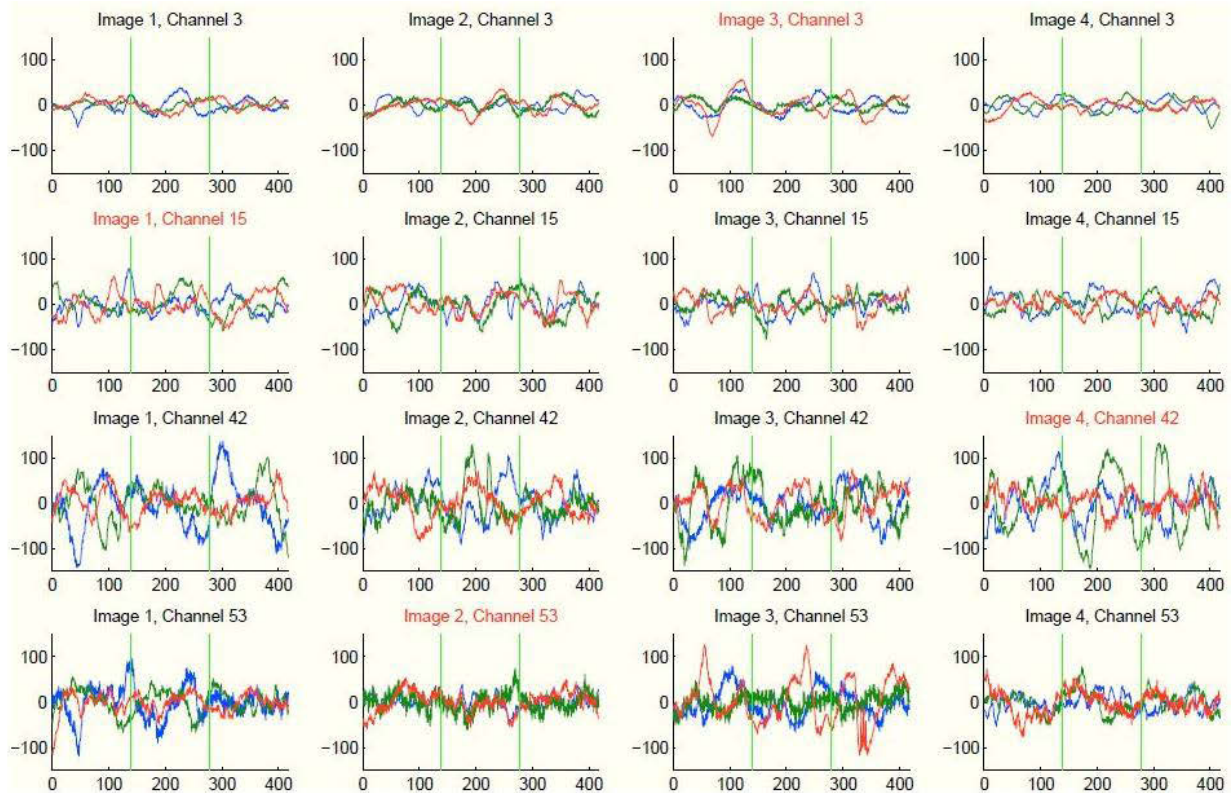


Figure 127 – LFP traces from a single patient

Each row corresponds to a channel. Each column corresponds to a different image. The channel that was sensitive to a specific image has a red colored title. Each subplot contains three mean LFP traces that were taken 1 sec before image onset (marked as green vertical line), up to 1 sec after image offset (marked as another vertical line). The three trace colors (blue, red, green) correspond to the mean trace in sessions 1, 2, 3. Mean was computed from all 12 available traces.

The other approach is using discrete time short time Fourier transform (STFT). The problem with STFT is its frequency resolution. A wider time window gives better frequency resolution but poor time localization. A narrow window gives good time resolution but poor frequency resolution. If we use a window of N samples and sampling frequency is f_s , then the Fourier transform gives $N/2$ coefficients (real signal). These correspond to $[0.. \frac{f_s}{2}]$ (i.e., Nyquist). Thus, the frequency resolution is given by (...) and can be approximated for large N as $f_s = N$.

I start my frequency analysis by asking whether there is a difference between the frequency range at the intervals $[i-1..i, i..j, j..j+1]$, where i represents image onset (in sec), and j represents image offset. I use the following notation: "before" refers to $[i-1..i]$, "during" refers to $[i..j]$, and "after" refers to $[j..j+1]$. Frequency estimations are obtained using *pwelch* function. This power spectrum density estimation method divides the interval into 8 intervals, applies hamming window, computes the DFT, and then averages over all eight windows to get the final spectrum estimation.

Figure 128 shows a nice example how power spectra in the hi-LFP range increases in the "during" interval, but returns to lower values at the "after". Interestingly, this phenomenon was observed only in this specific channel. In the three other channels no "clear" divergence of frequencies was observed (see Figure 129).

To test how significant this phenomenon is, we first applied one-way sample-by-sample ANOVA test to check whether the different intervals have a mutual mean or not. This is done per frequency. Since there are three groups and we use multiple comparisons (3 tests), we can

use a more stringent p-value for significance. In any case, the ANOVA results for channels 3 and 53 are shown in Figure 130 and Figure 131.

Difference between the different screenings

The most significant deviation of LFP frequencies from the "before" or "after" is demonstrated for channel 3 (Michael Jackson). Applying ANOVA on the "during" interval between the different screenings showed that there might be a significant difference (see Figure 132).

The power spectrum approach cannot pinpoint exactly when these frequencies changes occur. Since channel 3 seemed the most promising one we have focused our analysis only on that one. Figure 133 shows the STFT analysis averaged across image repetitions.

By pooling all image appearances from all screening sessions we generated averaged LFP amplitude (see Figure 134). However, the raw amplitude seems too noisy to be meaningful.

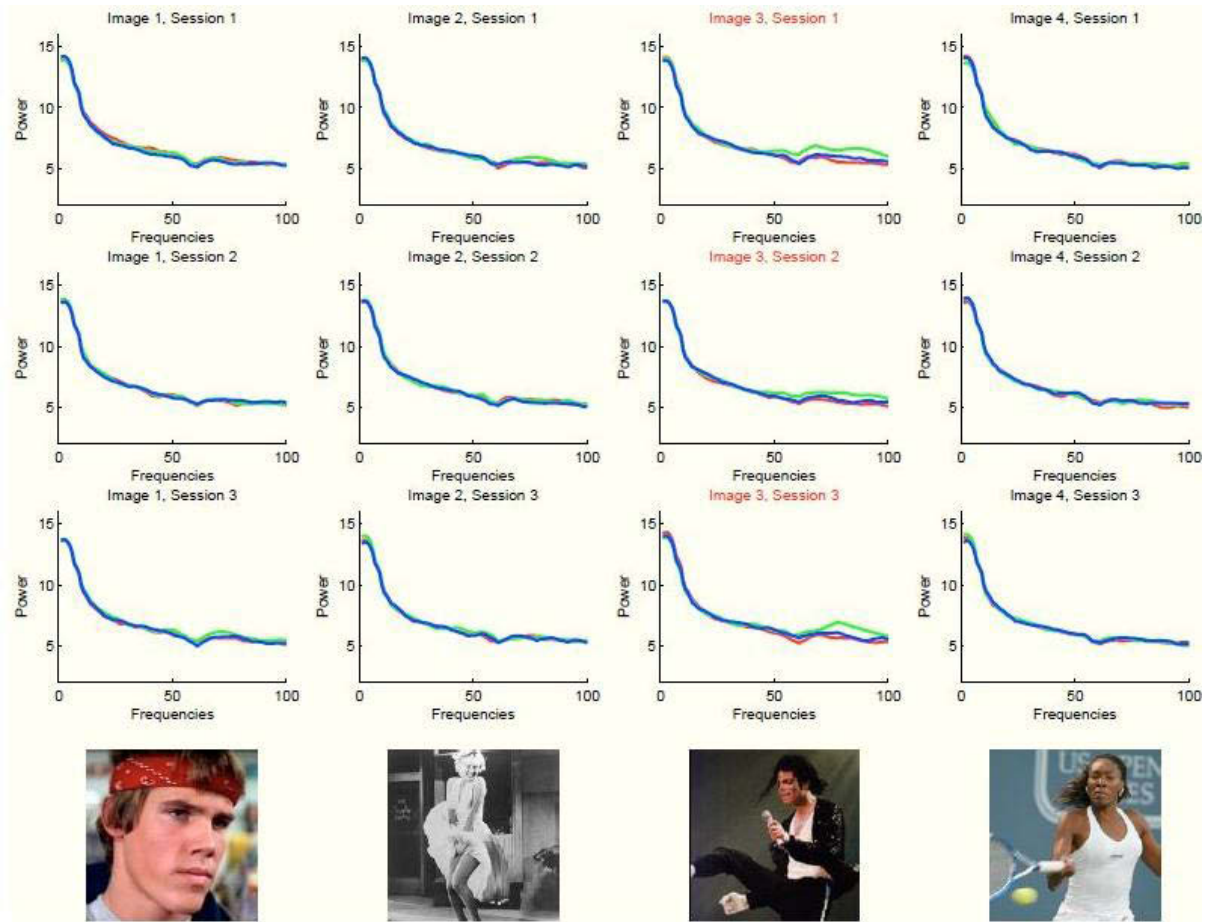


Figure 128 – Power spectrum density estimation

Spectrum is averaged over image repetitions. Red curve corresponds 1 sec before image onset. Blue curve represents 1 sec after image onset. Green represents spectra during image viewing. This data was taken from channel 3, which best responded to Michael Jackson image. Notice the significant increase in hi-LFP frequencies during image viewing, compared to the before and after intervals.

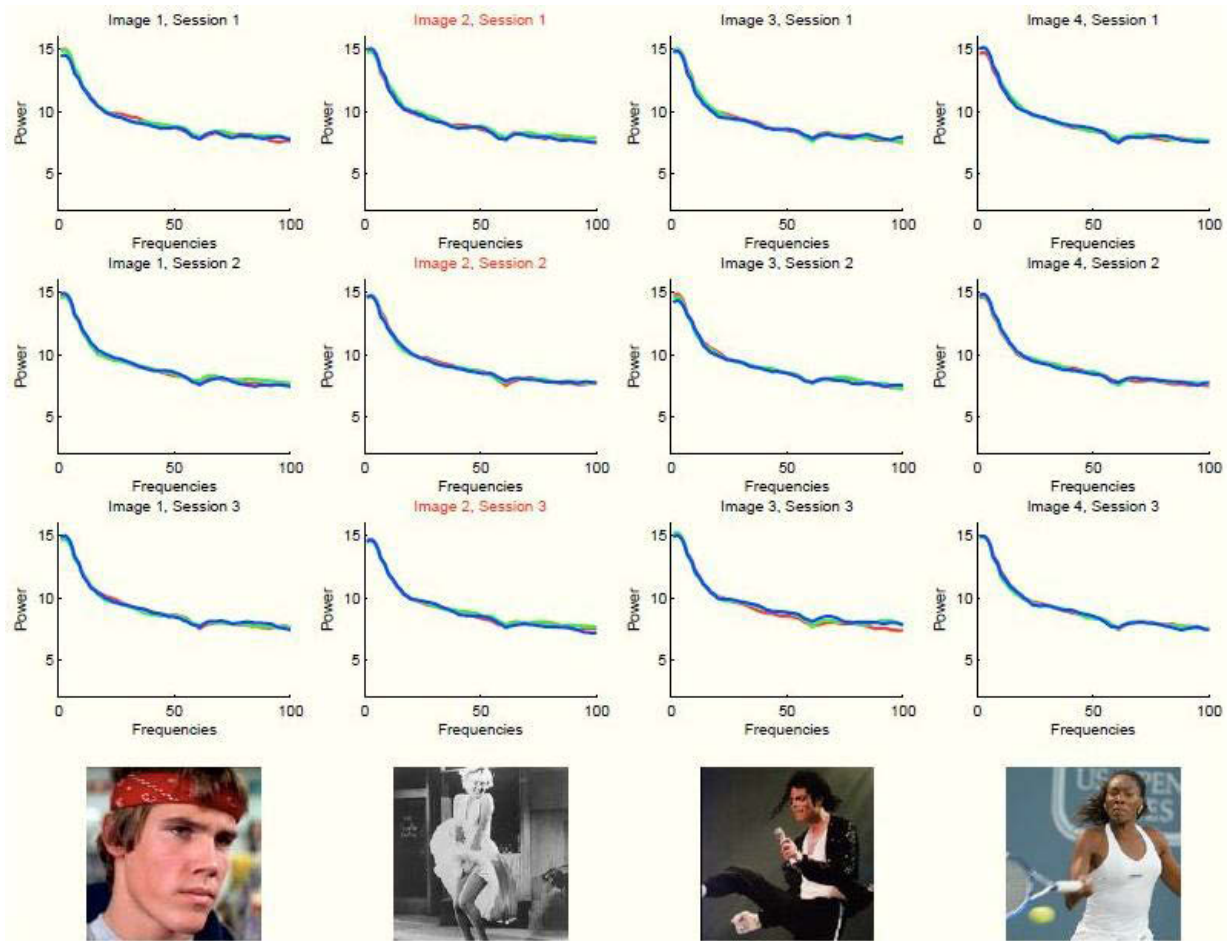


Figure 129 - Power spectrum density estimation from channel 53

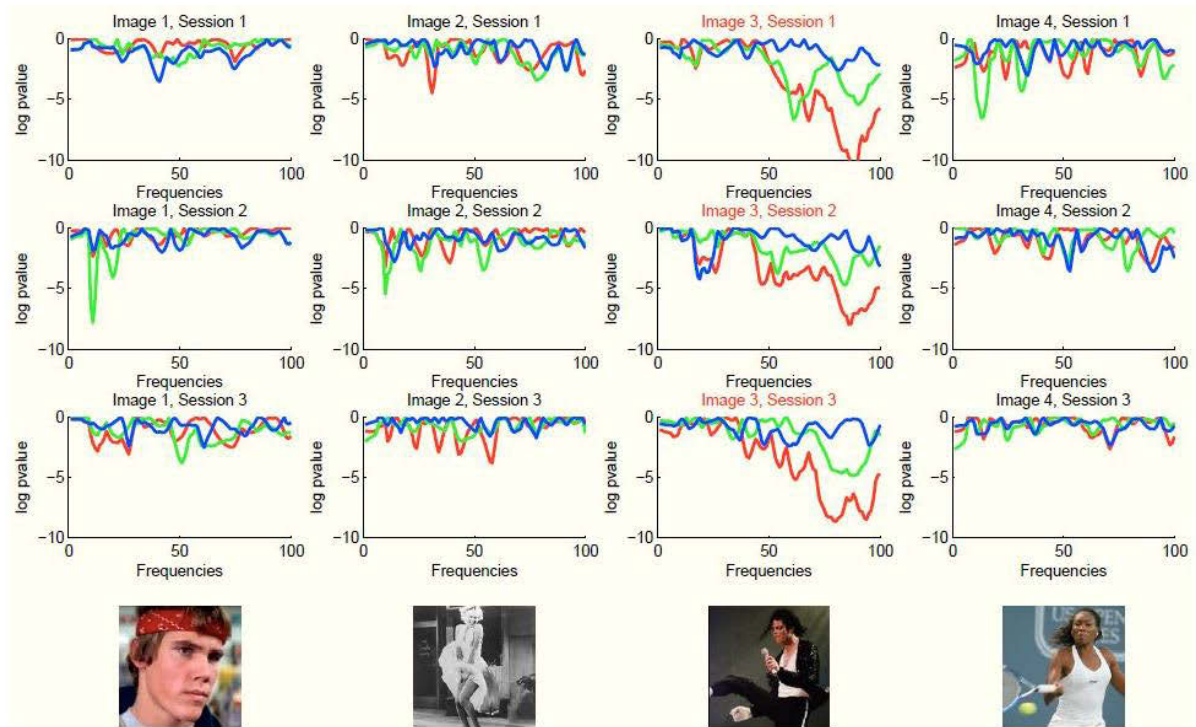


Figure 130 - One-way frequency by frequency analysis of variance for channel 3

X axis represents frequencies. Y label is log p value. Red curve represents the test between the "before" and "during". Green curve is the test between "during" and "after". Blue represents the test between "before" and "after".

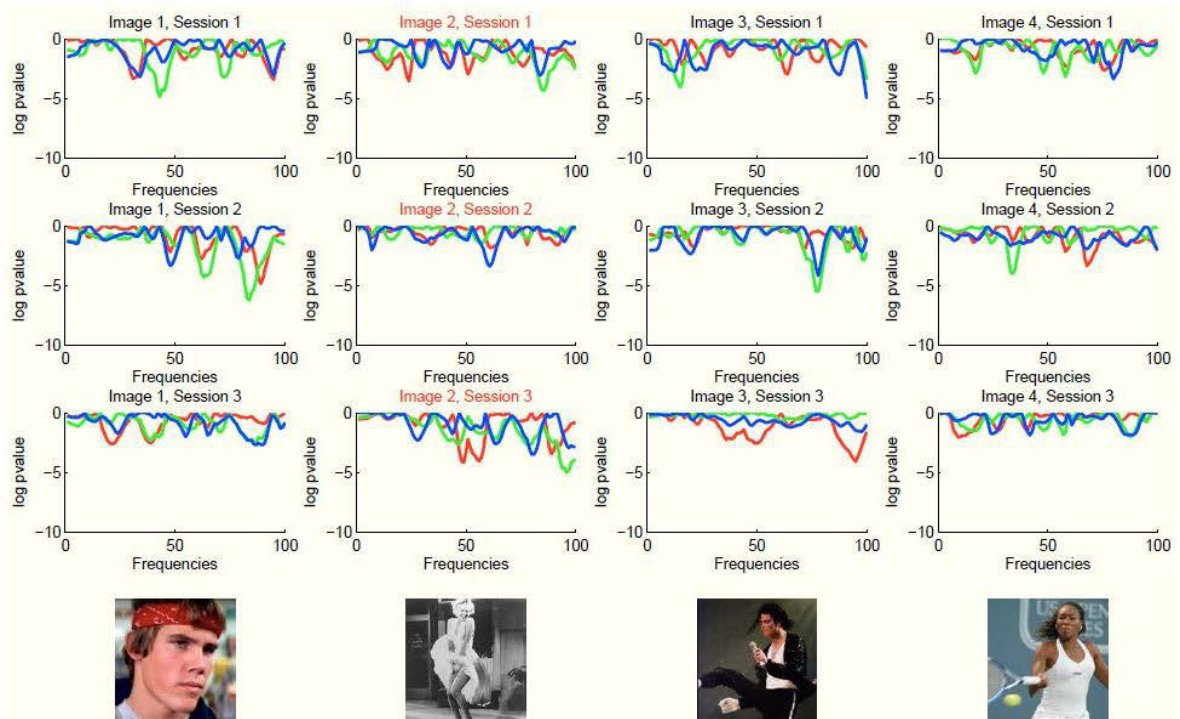


Figure 131 - One-way frequency analysis of variance for channel 53

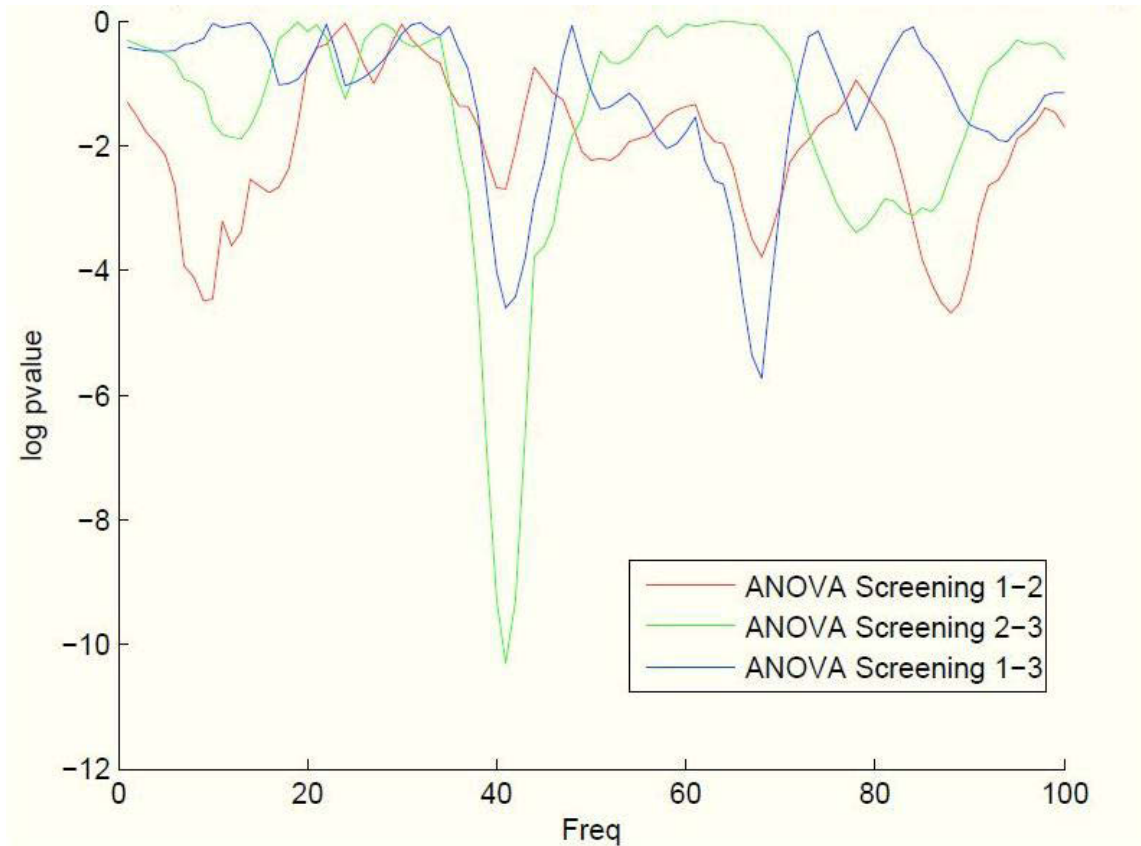


Figure 132 - ANOVA between different frequencies in the "during" interval

Red. Between screening 1 and 2. **Green.** Between screening 2 and 3. **Blue.** Between screening 1 and 3

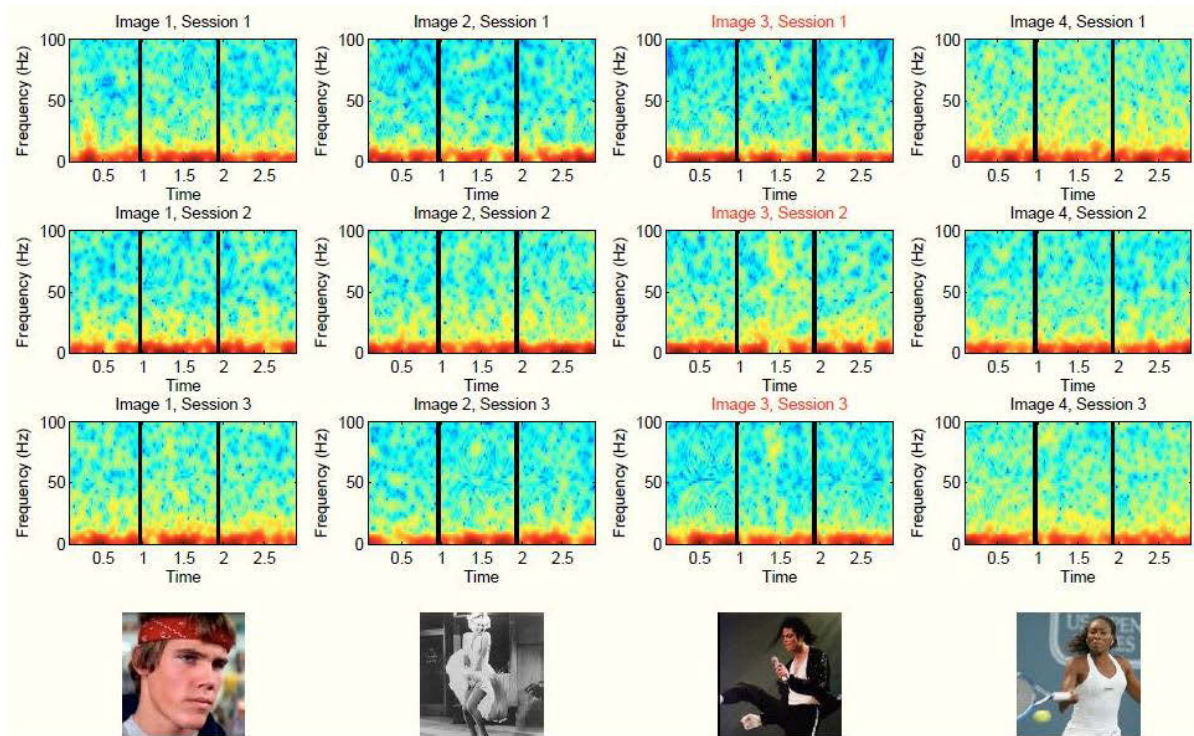


Figure 133 - STFT averaged across image repetitions

200 ms time window with 180 ms overlap was used

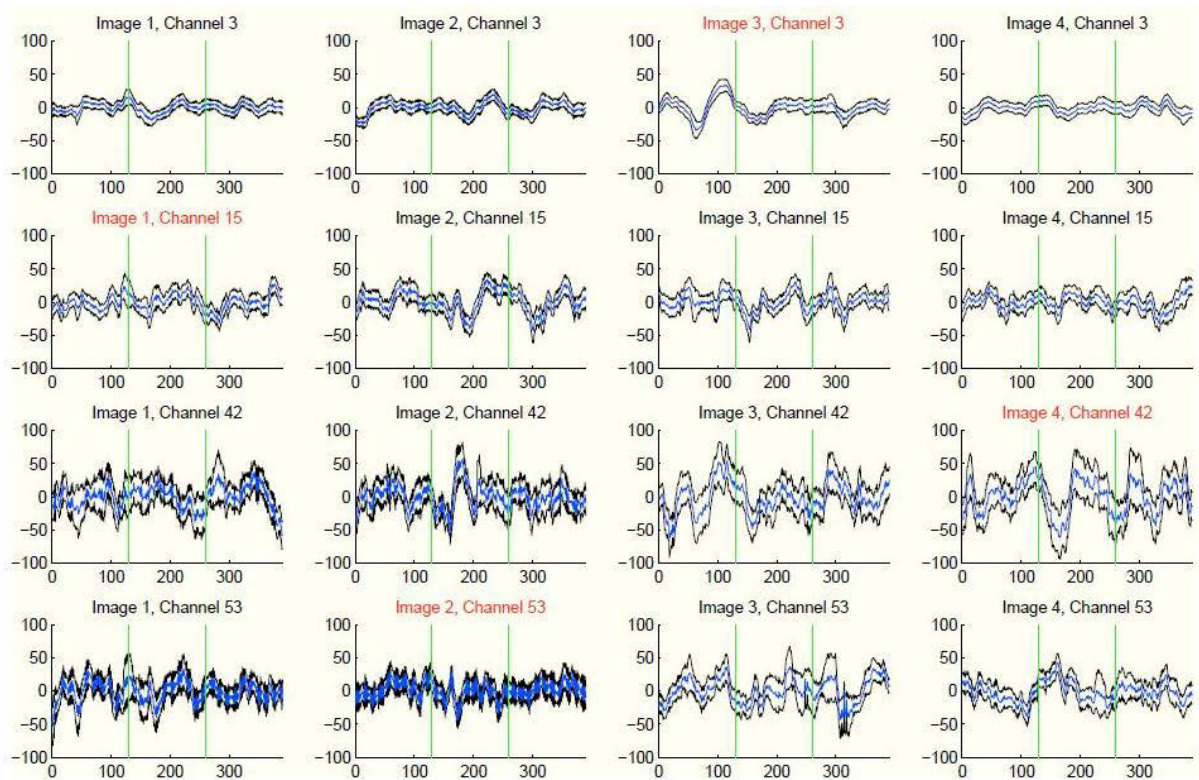


Figure 134 – Mean LFP traces and \pm S.E.M

It appears that several significant differences in frequencies can be observed. The most promising one is the increase in hi-LFP power. Whether there is a difference across screening sessions is yet to be determined. Figure 132 suggests a change in the 40 Hz or 70 Hz frequencies. The STFT analysis suggests an increase in the hi-LFP close to 300 ms after image onset.

Behavior

One anecdotal piece of evidence from the fading experiment with regards to attention is the ‘commitment’ of patients to making their best efforts in the game. As this experiment is solely conducted in the patients’ brains, and requires a great deal of effort on their behalf in trying to succeed in the task in the absence of clear instructions on how to do it, subjects may not put forth any effort or strain their attention and just fail continuously without trying.

Nevertheless, subjects tended to make an enormous amount of effort in trying to win the task and became very frustrated when they were not able to perform well above chance in it.

While the task is purely cognitive and requires absolutely no movement or any motoric involvement in it, we repeatedly see patients sigh in relief when the experiment ends and they can rest. A few subjects were indeed sweating during the experiment purely from the mental effort they were putting in performing the task well. More so, although clearly the control is only governed by the patients’ neuronal activity, we often see the patients’ patrons (parents, friends, or even our own experimenters in the room) taking effort in the supposed trial of the patient to perform the task. When the patient is close to getting the image to the target (say, getting it to around 80% of the target), but is struggling with the remainder movements, it is common to see the entire supportive crew focusing their thoughts altogether as if that has a value in assisting the patient (similar to the effort from a cheering crowd in a basketball game, although clearly the only players who can affect the game are on the field and not on the chairs outside of it).

As such, we commonly ask our patients to report their feelings regarding the experiment. Here are a few of their answers:

- “God. That’s SO cool.” **Patient 409.**
- “I see that I need to think of Jim, but I can’t avoid thinking of this horse.” **Patient 400.**
- “Damn. I’m good.” **Patient 400.**
- “I can easily get 3 out of the 4, but I just can’t get that David Grohl.” **Patient 409.** (Notice that in the context of a block it is impossible to get 3 out of 4. One can only get 0/2/4 out of 4, as 2 stimuli are exact repetition of the other 2 trials.)
- “My imagination doesn’t work too well today. The food here is awful.” **Patient 408.**
- “I keep fading to Hillary Clinton, and I don’t even like her. Why is my brain thinking of her? Now I am surely voting for Obama.” **Patient 402 (during the presidential primaries)**

Discussion

The past decade has seen major strides in the development of brain-machine interfaces on the basis of single neuron activity in motor cortex in both monkeys (Musallam, Corneil, Greger, Scherberger, & Andersen, 2004; Velliste, Perel, Spalding, Whitford, & Schwartz, 2008; Wessberg et al., 2000) and humans (Hochberg et al., 2006; Kim, Simeral, Hochberg, Donoghue, & Black, 2008). One difference to these studies is that we provide feedback from regions traditionally linked to memory acquisition and consolidation and emotional expression. It is possible that the rapidity and specificity of feedback control in our subject population is made possible by explicit and voluntary cognitive strategies that are directly matched to the capacity of these neurons to represent abstract concepts such as “Golden Gate Bridge” or “President Bill Clinton”. We previously estimated, using Bayesian reasoning, that any one specific concept, such as a familiar building or person, is represented by up to one million MTL neurons, but probably much less (Waydo, Kraskov, Quian Quiroga, Fried, & Koch, 2006). As our electrodes are randomly sampling a handful of units from the MTL, the cognitive control strategies of subjects must differentially modulate the firing properties of distinct groups of spatially interdigitated neurons.

Our experiment provides a look into the mechanisms that underlie the competition between external stimuli provided by the environment and internal mental states that govern thoughts. We are able to show that subjects can access areas deep inside their brains purely by focusing their thoughts, and altering the flow of information such that the internal states are overriding retinal-derived information. As subjects control the visibility of each image, generating trajectories of movement along a plane of representations by exciting or suppressing the activity

of individual neurons, our results suggest a possible analogy to equivalent place fields in the physical environment where the organism navigates. In an analogous manner, subjects may be seen as navigating a “concept space”. Our results make us optimistic about the possibilities of technologies to ‘read the mind’ of individuals unable to communicate their thoughts by externally projecting their thoughts on the way to a transhuman future.

Feedback in the vision system

The vision system is typically looked at as a feed-forward network. Information from the retina flows through various pathways until it reaches the medial temporal lobe where it is aggregated into an encoded concept (Van Essen et al., 1992).

Common studies suggest that a person is aware of approximately 10,000–30,000 concepts (based on the amount of words in the dictionary that people, on average, know). This complex diagram receives inputs from other brain regions on top of the complex connectivity demonstrated in the figure.

Inputs can come from various other regions and end up as concepts in the hippocampus. Prior studies have shown that the hippocampus can show similar responses to a concept even if the eyes are closed and the patient merely imagines seeing a visual stimulus from the retina (Kreiman et al., 2000b). This implies that parts of the same diagram can be activate in the absence of inputs from lower parts (for example, the hippocampus can be activated in the absence of activation of the retina). Our fading experiment demonstrates an even stronger level of abstraction of the inputs. The patient can see one stimulus – activating all the channels in the diagrams corresponding to activity from that stimulus – yet imagine or focus his thoughts on a totally different stimulus that exists purely in his imagination at that point, and still override the information – making the hippocampus, for example, respond in a manner that seems as if the information from the retina never propagated in the visual system. This suggests that either inputs from external mechanisms to the ones shown in the Van Essen diagram are governing the attention, or that feedback can indeed alter this network in a way that makes it not act as a feed-forward one anymore, but rather feed-back. This concept could suggest a very different view of the flow of information in the visual system, suggesting that we can manipulate the activity in various areas in it that were regarded beforehand as inaccessible. Although we know what part of the visual system, for instance, is responsible for processing color information in images we see, we cannot voluntarily decide to shut off this module in V1 and

start seeing in black and white – we simply can't access this region in our brain directly. The feedback experiment demonstrated here suggests that some deeper levels of our system that are regarded inaccessible are in fact accessible to us given a proper mechanism that tweaks the flow of information through a feedback mechanism that rewards the network for behaving in a manner that supports changes in flow of information, or changes in weighting of the competing channels towards a single concept perception.



Modeling Attention

*I've always seen modeling as a
stepping stone.*

Tyra Banks

In science, if you really want to have a solid explanation of a phenomenon, in my opinion, it is not enough to test it and see that the results make sense and are significant. The ultimate explanation is a model that can make predictions and be verified with results you have not yet seen from your data. This is what we tried to do with the data from the fading experiment. We took the data and built a computer model that fits the results of each and every patient separately and then varies the level of attention allocated by this particular patient to tell if we can learn from his behavior something about the nature of attention in general. This chapter describes the model and its exciting output.

Method

We modeled the attention of each patient using a neural network comprising of four simulated neurons, representing the ones recorded from during the fading experiment. Each neuron had a firing rate λ_0 at baseline (i.e., 1 Hz for neuron 1, 0.1 Hz for neuron 2, etc.) and an increased firing rate λ_i when its preferred stimulus (different and unique for every neuron) is presented. The probabilities of alternating between the baseline state to the active state are γ_1 and γ_2 , accordingly. These probabilities change based on the presented stimulus (i.e., when a neuron's preferred stimulus is presented the probability γ_1 is higher (more likely to become "active") and the probability γ_2 is lower (less likely to return to baseline), while those are reversed (more likely to stay in baseline than become active) when a different stimulus is presented.

At first, a random screening is created, with the neurons' firing rates and probabilities fixed, to create a decoding matrix corresponding to the presentation of the stimuli. The matrix is created in the same way the experiment is produced. That is, each neuron's firing rates are calculated for the course of the entire experiment based on the order of images presentations. After the experiment ends, each of the 48 seconds of images viewing (12 repetitions x 4 images x 1 second) are binned to 10 100 ms, and the 7 last ones (300 ms to 1000 ms) are placed in a matrix where each bin's firing rate is labeled with the correct stimulus that was presented at that time. This yields a 4 (neurons) x 336 (7 bins' firing rates x 48 images) numbers corresponding to each neurons' activity during each of the 48 images presented.

Fixing the decoder matrix we simulated 16 fading trials for 2 different preferred stimuli (8 trials each). At each trial a target is randomly selected and presented for 2 seconds to the 4 neurons. The probabilities γ_1 and γ_2 for the preferred neuron are increased by an "attention" parameter,

making the likelihood of having the neuron active higher, while the probability of going back to baseline is decreased. The state-change probabilities for the 3 neurons that are not affected by the “attention to the preferred target” remain fixed. For the 2 neurons that see a visual faded feedback of the preferred stimulus and its distractor, the probabilities to increase the firing rate based on the visual content (as done in the screening) are now multiplied by the % of the image that is seen.

For example:

Say in trial 2, the target is image **B**, and chances of changing from “baseline” to “active” for that image is $\gamma_1 = 0.4$ for the neuron whose preferred stimulus is **B**. Then, the assumption that the patient is thinking/attending to the given stimulus is viewed as an increase of, say, 0.2 in the chances of changing to “active” state — making γ_1 now be $0.4 + 0.1 = 0.5$. For neuron 2, whose preferred stimulus is, say, **A** (distractor in this trial), there is no attention to the stimulus (the patient is thinking of **B**, not of **A**), hence the chance of changing from “baseline” to “active” isn’t affected by the attention. If the initial probability was $\gamma_1 = 0.3$ for that neuron, then it will remain the same. However, since during the first 100 ms of the trial the patient sees a “morph” of the two images (**A** and **B**) the visual feedback affects both neurons. The effect of the feedback is therefore modeled as a multiplier of the amount of visual cue on-screen (in the beginning 50% for image **A** and 50% for image **B**) by the maximal probability of firing in case the image is fully presented on the screen. That is, if the chance of becoming active for neuron 1 is $\gamma_1 = 0.6$ in the screening, and the visual feedback at the first 100 ms is 50% than the increase in probability for that neuron to fire is $0.6 \times 0.5 = 0.3$. This number is added to the attention and baseline probabilities, such that: $\gamma_1 = 0.4 + 0.1 + 0.6 \times 0.5 = 0.8$. That is the

chance of making the neuron fire in that 100 ms bin. For the other neuron (not preferred), the probability will be $\gamma_1 = 0.3 + 0 + 0.7 * 0.5 = 0.65$ (if the probability of changing to active in the screening was 0.7). The probabilities of the neurons that were not attended, and had no visual feedback (neurons 3 and 4) remain fixed.

All neurons' spikes at the 100 ms bin are then summed to give a population vector which, in turn, is compared to the pre-labeled decoder to reach a decision regarding the likelihood of that 100 ms having a belonging to any of the 4 stimulus. Based on the results, a step in one of 3 directions is performed. If the decoder identified the population vector as closest to stimulus **A**, the image will advance 5% towards **A**, if it is decoded as **B** it will advance 5% towards **B** (e.g., from 45/55 to 40/60), and if the decoded resulted in either **C** or **D**, the image fading won't change.

Following is a list of the parameters and variable of the model:

- Neuron: $i \in 1..4$
- Activity: λ (Hz)
- Probability of changing state: γ

Which is a function of the following two entities:

1. Feedback $\in 0..1$ (showing the amount of fading between the target image and the lure)
2. Attention $\in 0..1$ (the variable measured, showing the increase in attention to the target in a given trial)

Table 10 - Parameters and variables used in the fading attention model

See Figure 136 for an illustration of the neural network.

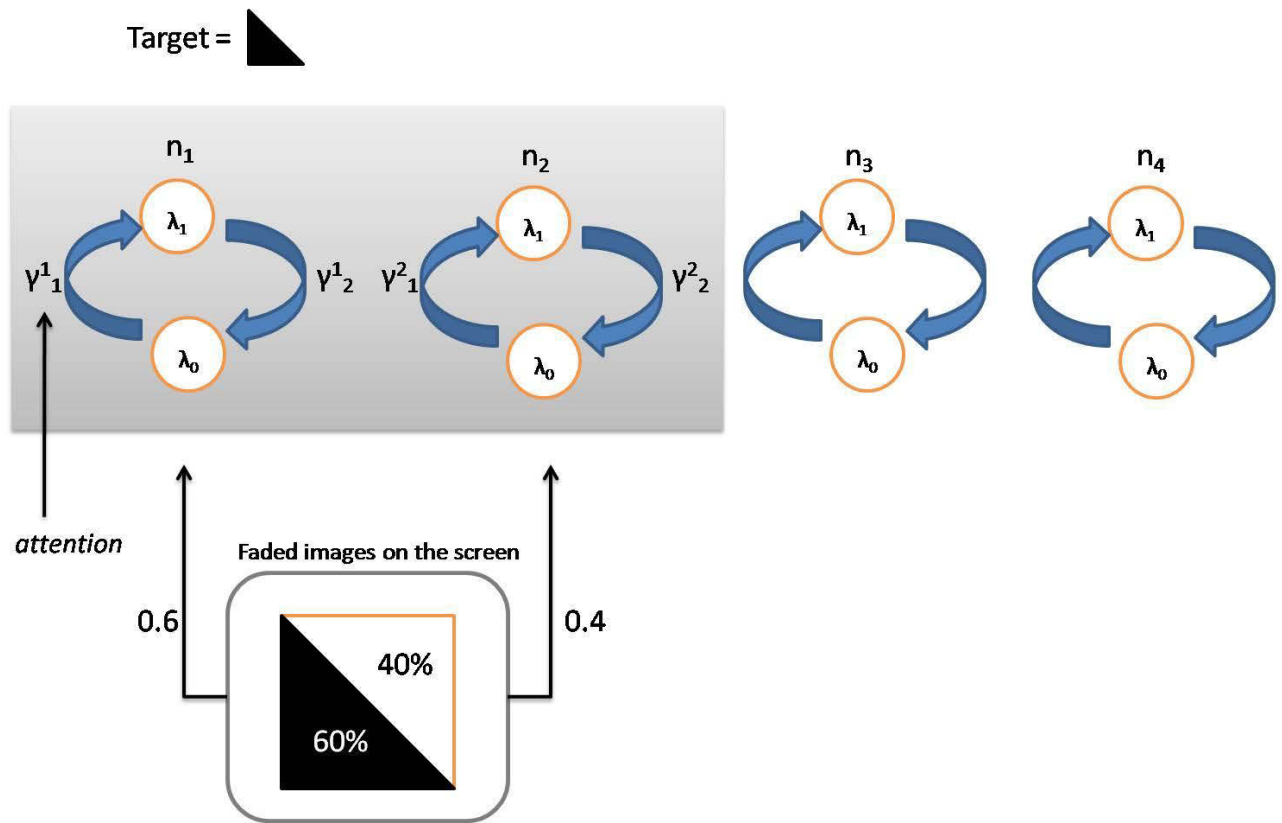


Figure 136 - An illustration of the computer model

In the illustration, for the shown trial, the target is illustrated as a black triangle. On the screen the patient is showing 60% of the target and 40% of the lure (white triangle). This, in turn, adds 0.6 to the probability of neuron 1 (n_1) changing its state from "baseline" (λ_0) to "active" (λ_1), and adds 0.4 to the probability of neuron 2 (n_2) as some of the feedback is of the lure. Neurons n_3 and n_4 are not affected by the feedback, but are still able to change their state to "active" under some (low) probability. Neuron 1 also receives additional boost from the attention, which varies from 0 (no attention to the black triangle stimulus) to 1 (full attention to the stimulus).

The current model doesn't account for the following parameters, which are, thus, not part of the modeling:

- No effect of training – trial i is equal to trial j .
- No inhibition – γ of neuron i doesn't affect γ of neuron j .
- No plasticity – λ_0 and λ_1 do not change during the experiment.
- No latency – The neurons are all “ticking” at the same clock. No neuron fires before the others.
 - No “starting thought” due to the target viewing before the trial – All neurons start changing their states at $t = 0$, and no neuron starts with shorter latency (the “tick” is 100 ms for all neurons).
- No “regional” differences – Neurons' baseline/active firing rates are all the same

Results

We separately fit each patient's data in the following way: we first tested the number of spikes per second during baseline activity while the subject was performing the initial screening prior to the fading block. For the "active" state we took the firing rate (Hz) of each neuron during the times when its preferred stimulus was presented. These parameters (λ_0 and λ_1) for each neuron and each patient were fixed from then on.

We then counted the number of times the neuron changed its state from baseline to active during the presentation of the preferred stimulus, in each 100 ms (out of 10) during the 1 second image presentation. This gave us the parameters γ_1 and γ_2 by the proportion of times the neuron was in the active mode during image presentation (e.g., if each time the preferred stimulus was presented the neuron was indeed in the active state, then γ_1 would be 1.0 and γ_2 would be 0.0. If 5 out of 6 times the image was presented the neuron was in its active state then γ_1 will be 0.83). These parameters (γ_1 and γ_2) too, are fixed from then on for the course of the study, for each patient.

Finally, the fading is updated every 100 ms from the starting of each trial, based on the amount of fading each image has (starting at 0.5 and advancing in 0.05 steps towards 1 or 0), and the only free variable in the study is the attention which varies from 0 to 1 for each run.

A run consists of 32 trials with a fixed attention number yielding a measure of both performance and time: for each of 32 trials we could tell how many of them succeeded and failed for that particular attention level, and we could also measure the timing of the trials. Using those two numbers we selected (mean squared error) the attention level that best fit the

results of the patient. This is regarded as the attention level that the patient actually exhibited in the course of the experiment.

Clearly, finer metrics could be used and more detailed measure of the attention could be applied. For instance, one could fit not the entire 32 trials as a block, but each trial separately – resulting in 32 different attention numbers. Then the average of these would be regarded as the attention level utilized by the patient. However, our metric showed significant results and is more general – assuming that a patients internal attention “metric” is consistent and does not change with time.

We repeated the attention fitting with the given parameters 1000 times for each patient, resulting in varying levels of attention based on the results (the only variable that is free, within each attention number, is the random number chosen in each 100 ms update that decides for each of the neurons the state within which they are at).

For ten patients and 1000 trials, we estimate the attention contribution to the change in the neuronal response to be 0.21 on average. See Figure 137 for the attention estimate for each patient.

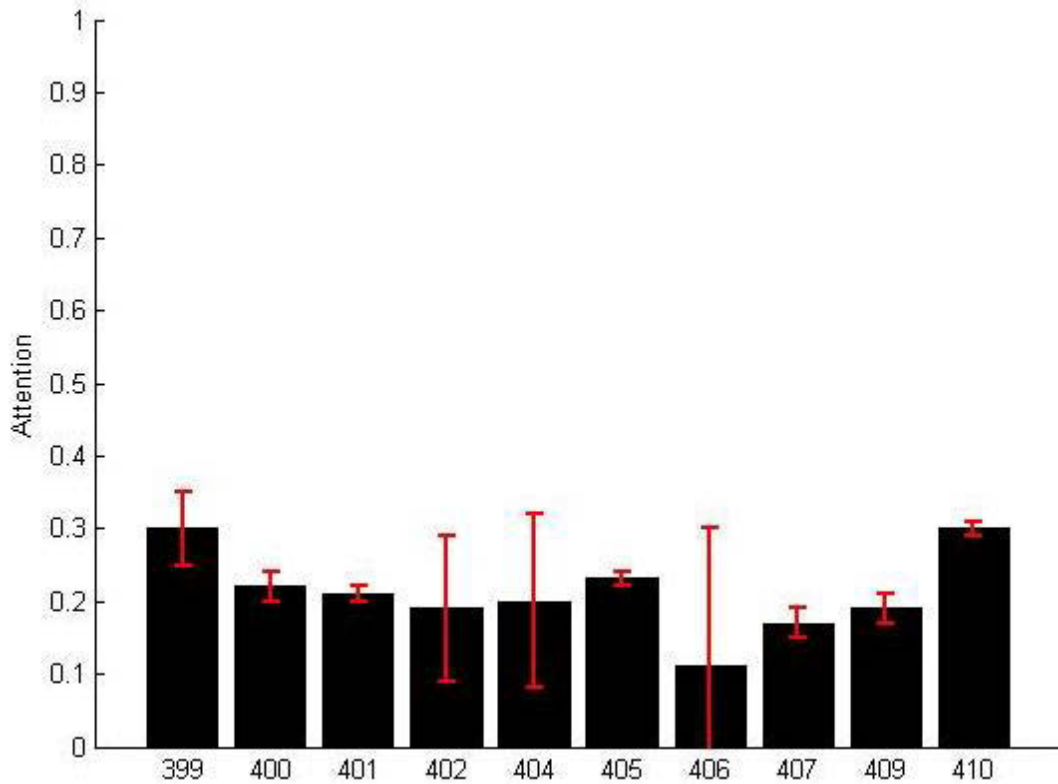


Figure 137 - Estimation of the attention level for each patient

Each bar represents the averaged attention out of 1000 trials. Red lines are error bars.

The fitting error was never larger than 0.3. If the error was bigger the entire trial was discarded as this indicates a poor fit of the data to the patient. On average the errors are below 0.05, indicating a strong fit, and as such are a good estimate of the attention given the fixed parameters.

Discussion

We showed a model representation of four selective neurons in the human brain yielding results similar to those reached from ten patients in an experiment. While it is unclear what mechanisms in the brain govern the attention, and it is hard to claim that these are singleton mechanisms, or that they are identical in each patient, we can show here that the average level of attention is similar across ten independent subjects. Testing for the variability across subjects shows significant results only for a single patient (406), while all others are indistinguishable, suggesting that the attention is exhibited in a similar fashion for all patients in our model.

While attention reaching a level closer to 1.0 would make any other external input, such as the feedback, useless, as the majority of the contribution will be rising from the attention itself, we can expect to see much lower levels of attention as the proportion of contribution of attention to the control is still very high (a third of the effect of feedback in most patient can be explained purely by the attention modeled by top-down decision of the patient to focus his thoughts on a the target and avoid the lure).

This model does not tell us what constitutes the remaining 70% of control over the fading, but between the feedback from the images, and the internal probability of the neurons' state changing we can estimate attention as having the majority of control over the fading process.

This model, thus, suggests a simple and unique way of breaking apart attention and external inputs from the environment and allowing for a quantitative appraisal of the contribution of each. If we look at our brain as a black box that receives inputs from the environment, processes them and acts upon them, then this model can show that the majority of the decision making is based on the processing and not on the external input. Or, to put it differently,

subjects who would have performed the task entirely with their eyes closed and controlled the fading with their imagination (pure attention-based control) would do much better than subjects who would not know what the target was and would just passively follow the fading based on the visual feedback (pure feedback-based).

This model, therefore, is a unique approach to disentangling attention from perception and serves as a suggestive addition to the data shown in the previous chapter.



Computer Games

But why is it fun?

Randy Pausch, in his "Last Lecture" describing the guideline to his every new experiment

What a great way to end my thesis: with a game. Everyone likes playing, and computer games are one of the most engaging, exciting, and entertaining way to teach a subject, get people to work on a subject, or gage attention. Naturally, when we designed the experiments with our patients – wanting them to pay as much attention as possible to our task – finding a cool computer game that would make them actually “play” rather than be subjects in a study was a highlighted goal. It took us over a year to actually see the first patient enjoying the game, but when it actually happened – boy, was he thrilled. The goal behind the game was that, in the rare case where our MTL patients ended up having perfect or significantly high level of control over a single neuron in their brain, we would have them control “something” with their mind alone. This “thing” would not be a trial-by-trial task, but actually a continuous game. Our initial challenge was that the game had to be such that it could be played with only one binary neuron. The patient can turn his neuron “on” (increase the firing rate above a certain threshold), or turn it “off” (by having the firing rate be below the threshold). Later on we improved the game a little by making the scale continuous and having the patient in fact control the threshold itself – still having control of only one neuron, but making the game a “one-button game”. As excited as we were about the ability to actually design a computer game, we immediately realized the challenge of designing a game that works with only a single button. Wanting our game to be exciting (some children games are “one-button games”, but they would most likely not entertain a 40-years-old patient) and at the same time simple and one-button proved not to be as hard a challenge as we had imagined. Apparently, there’s a large community of geeky teenagers across the world who vividly study and design such games. Samy Zerrade – a SURF student I had the pleasure to work with during my third year at Caltech –

was the one to find an entire website titled “one button games”; from there the road to our own little “space invaders” was very short.

Introduction

Being able to read humans' minds and translate thoughts into control of physical objects in the real world has been a subject for science fiction stories since long ago. Advances in the field of brain machine interfaces (BMI) have recently gotten us closer to achieving some of those seemingly imaginary tasks (Carmena et al., 2003; Hochberg et al., 2006; Musallam, Andersen, Corneil, Greger, & Scherberger, 2005; A. Schwartz, Cui, Weber, & Moran, 2006; Velliste et al., 2008). From the very early suggestions by (Fetz, 1969) to the more recent examples shown with animals (Velliste et al., 2008) and with paraplegic patients (Hochberg et al., 2006), it is clear that being able to directly read intentions or facilitation of actions in the brain's motor area should unveil an endless amount of possibilities both to help the scientific community better understand the brain and its coding of actions, and to assist paralyzed people and people with physical disabilities. These latter populations would potentially benefit from these advances to the point where we can imagine a stage where they would be able to reach a level of control of a prosthetic device that would allow them to regain operational usage of their amputated limbs.

While there are some suggestions of non-invasive devices that could serve as an interface between the brain and the computer (using EEG, MEG, fMRI, etc.) it seems natural to believe that a more intuitive method of brain control would be by applying direct signal from the neurons encoding information in the brain. However, these types of studies are scarce because of the rarity of direct accessibility to humans' brains.

Previous studies of patients with electrodes implanted in their brain have shown neurons selective to pictures of famous people, landmarks, animals, and various categories such as

shapes, household objects, cars, and more (Kreiman et al., 2000b). Further studies of similar patients have shown similar neurons to be invariant to the content of the image – responding selectively to concepts (Quiroga et al., 2005). These neurons, thus, will increase their firing rate significantly when presented with various different images of a person, for instance. More so, further studies have shown that these neurons will occasionally also fire in a similar fashion when the patients close their eyes and merely imagine a concept from a certain category (Kreiman et al., 2000a). These suggest a very unique coding of the concept by those neurons. A coding that may well be suited for precise voluntary control of an external entity, if a patient could voluntarily trigger the responses elicited by these neurons in the presence of his own mind.

In this study we recorded directly from the brains of two human patients undergoing brain surgery for clinical reasons, while they performed a task where they were to consciously and voluntarily elicit activity in a single neuron in their brains, trying to play a computer game with their mind alone. We tested the ability to gain access to areas of the brain that are commonly regarded part of the sensory system, and as such not typically directly accessed voluntarily. We assessed the performance of patients playing the game, as well as the ability to do so in a short time with very little training.

In order to have the patient play the game we developed a system that detects activity in patients' brains, and using a real-time detection system decodes the activity in the brain and performs an action accordingly, in a manner that seems to the patients as directly influenced by their thoughts. This system is scalable to between one and many neurons' signal detection in

real-time, and uses an algorithm that allows for complex decoding of thoughts the patient is raising.

The contributions of this study are: (1) Experimental data showing that patients exhibit significantly high level of control of a computer game using a single neuron in their brain, (2) quantitative results showing application of a system that is commonly looked at as sensory to a willful voluntary motor-like control, (3) results showing a very short time (minutes) to reach this high performance compared to previous trials with humans and monkeys that lasted weeks, months, or years, (4) real-time testing of a novel system and algorithm for detection of spikes and decoding of patients' thoughts given prior knowledge of selective activity in their brains.

Methods

Experiment

We directly studied the ability to form a brain-machine interface using data recorded in real-time from single neurons in the human brain. Subjects were two patients (Patient 1: male, 31 years old, right handed; Patient 2: female, 42 years old, right handed) with pharmacologically intractable epilepsy implanted with chronic electrodes to localize the seizure foci for possible surgical resection (Fried et al., 1997). Based on clinical criteria, electrodes were implanted bilaterally in the amygdala, entorhinal cortex, hippocampus, and parahippocampal gyrus. Studies began by running a 30-minute screening session where approximately 100 images were presented to the patients and s/he was asked to look at them for exactly one second and respond as to whether they included an animal or not. 64 channels were recorded using a Neuralynx system (Neuralynx, Bozeman, Montana). An analysis of the responses of the channels during the screening session revealed selective neurons in both patients. Out of these, a few images were chosen as candidates for further analysis, based on locations in the brain and significance of the response (see (Quiroga et al., 2005) for selection criteria). After various testing with the patients during the day, we selected the most consistent responses that remained significant in various different tests as candidates for the real-time feedback computer game experiment.

The game

The game played by the patient is a simplified version of a one-button obstacle course. The screen shows an airplane on the left-hand side flying right. At random times an obstacle would

show up from the right and would slowly advance towards the airplane. The task of the patient is to increase the height of the airplane in order to avoid colliding with the buildings (see Figure 139 for an illustration of the game). Buildings vary in height. The height of the airplane's flight level is regulated by the firing rate of the selective neuron used. That is, if the baseline firing rate (calculated prior to the game) of the neuron was 0.2 spikes/100 ms (5 Hz), we would regard that as the airplane being on the ground, while any activity above that threshold would increase the height of the plane. The screen was segmented such that the maximal height would correspond to 5 standard deviations of the baseline activity. Buildings' maximal height corresponded to 4 standard deviations from the baseline. Thus, patients were asked to activate the neuron 5 times more than its baseline in order to avoid colliding with the tallest building. The game is played for 3 minutes, with an obstacle layout that is generated randomly prior to the game. We tested the ability of patients to control the airplane in the obstacle course above chance using activity from a single neuron in their brain. Patients were told ahead of time what concept drives the activity of the neuron and were trained for 22 minutes (patient 1), and 41 minutes (patient 2) in various other feedback tasks in controlling the activity of the neuron based on acoustic feedback indicating the amount of activity (firing rate) elicited by their ability to form a mental imagery of the concept. Patients were asked to keep their eyes open during the game (as we saw earlier that subjects perform the mental imagery better with eyes closed). The obstacle course was set such that no two obstacles will appear in a period shorter than 3 seconds apart, and the screen allowed for viewing of each obstacle 7 seconds before it would hit the airplane, giving the patient time to prepare their mental imagery. Patients were very engaged in the task, and reported being exciting about playing it as well as very eager to succeed in it.

Configuration

A system comprised of 3 computers was established in order to facilitate the analysis and presentation of the computer game in real-time (Figure 138). Eight wires containing data for eight micro-electrodes each, originating at the patient's brains, were bundled and referenced upon connecting to an acquisition system. The system was set as a server and as such sent data for all 64 channels. Data was stored in files locally for future offline analysis. A remote client computer connected via TCP/IP to the acquisition system received the raw data from the relevant channels. The network was set up such that no latencies would allow for delays in the data transport. The initial client system ran a C++ code (<http://www.klab.caltech.edu/~moran/feedback>) that ran a faster and enhanced version of the Matlab code given in (Quiroga, Nadasdy, & Ben-Shaul, 2004) for spikes detection. This code has a built-in latency of 2 ms for a single channel (size of the buffer allocated for data handling of the acquisition system), and scales logarithmically upon an addition of multiple channels. Raw data is band-pass filtered, and a threshold is applied in order to detect the spikes. Spike counts within a 100 ms window are calculated and forwarded via TCP/IP to the experiment computer. On the experiment computer the firing rate for the last 100 ms is added to a queue of 7 bins which are summed and displayed on the screen. A sliding window moves every 100 ms, including the recent entry. This smooths the activity of the “airplane” (the triangle the patient controls) making it flicker less, which is a concern when running experiments with epileptic patient (Benbadis, Gerson, Harvey, & Luders, 1996).

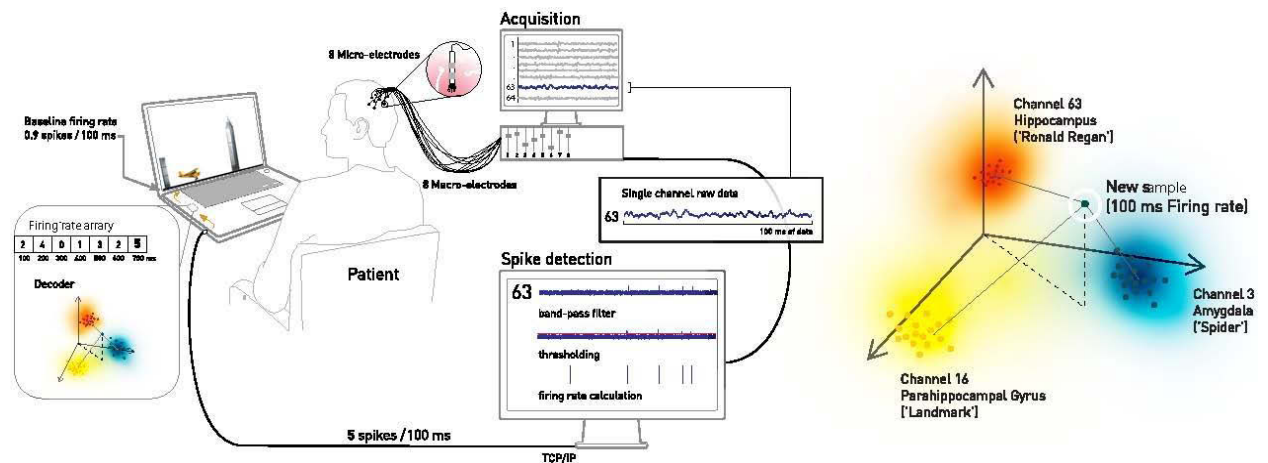


Figure 138 – Illustration of the configuration used in the experiment

The patient sits in front of a laptop display showing a computer game. 64 micro-electrodes implanted in various areas in the medial temporal lobe are bundled into 8 macro-electrodes (8 micro per macro) that are connected to the acquisition system. Analog raw data from the patient's brain is sent to the system where an A/D conversion of the data occurs. Data is then stored in the memory and immediately duplicated to a wire sending a selected number of channel's raw data to the spike-detection machine. The data is sent in packets of 100 ms. The number of selected channels can vary from 1 to 64 based on the experimental needs. In our case we always used a single channel as the input for the discussed game. Signal at the spike-detection machine was band-pass filtered from 300 to 1000 Hz, then thresholded according to the criteria discussed in (Quiroga et al., 2004). Events above the threshold are regarded as spike. For each spike we keep the time of the peak, and 32 points before and after in order to store the spike's shape for future clustering. The single integer firing rate is sent over a TCP/IP channel to the experiment computer running the game. Firing rate are added to an array where a sliding window of 7 cells that advances by 100 ms every 100 ms is averaged to come up with the next value to send to the decoder. Values were averaged across 700 ms bins in order to

refrain from creating too rapid changes that have the unlikely possibility of eliciting a seizure within the patient. Finally the decoder matches the number to a pre-defined space of stimuli optional entities to which it assign a decision value. In the game's case there was only one entity for the decision and as such the decoder just reflected the firing rate directly as an increase in the airplane hover height. In a more complex task (not reported here) a decision based on population vector could determine various optional maneuvers based on the decoding of activity of various neurons. The system is scalable to large number of neurons with an estimated delay of 7 ms per additional channel to analyze. Finally, the airplane is changing its height directly based on the activity of the patient's single neuron corresponding to his thinking of a concept that was already determined to have this neuron be selective for – and defacto is controlling the height of the airplane with his thoughts with a delay of up to 100 ms. **Right panel.** An illustration of the decoding process for a space of 3 channels selectively responding for 3 different stimuli. A new 100 ms sample of firing rate for the 3 neurons would be regarded as a point in the 3D space. A classifier trained prior to the experiment with the responses for each of the stimuli and their potential activity pattern (during a repeated screening stage, for instance) will now find the distances to each of the clusters using the Mahalobi's distances and would determine which of the stimuli was thought of and accordingly which action to take.

The decoder – not used with the patients mentioned in this chapter, as it is single-channel one – is trained prior to the experiment using labeled data (e.g., the early screening session where the 2 relevant stimuli are presented to the patient). The firing rate during the training period is recorded in a hyperspace corresponding to the number of channels (see Figure 138 for illustration).



Figure 139 – Illustration of the computer game played in the experiment

Patients saw an airplane hovering over the x axis. Buildings are approaching from the right — suggesting a feeling of flying the airplane right. The patient's task is to avoid colliding with the building. He does that by controlling the height of the airplane by thinking of a given concept that he was earlier trained to understand is controlled by his mental activity. The buildings advance slightly (10 pixels) every 100 ms, giving the patient approximately 7 seconds to prepare for the appearance of each building. At each 100 ms we evaluate the success of the patient using 4 complementary values. True Positive (“Hit”) is a case where the patient was able to avoid colliding with the building (buildings' heights vary throughout the game, but are always below the maximal firing rate exhibited during the previous screening session, and above the baseline firing rate of the neuron); True Negative is a case where there was no building and indeed that patient didn't make the airplane increase its height above the baseline; False Positive is case where, in the absence of a building, the patient did increase the activity of the neuron above the baseline (this might also happen due to spontaneous activity of the neuron,

but is still regarded as a FP, as the patient can learn to inhibit the neuron, and in order to maintain a very restrictive approach where only at the time of the collision should the patient make the airplane hover; False Negatives (“Miss”) are collisions between the airplane and the buildings. The illustration demonstrates a 10 s sampled screenshot of the game showing the 4 performance evaluation variables used. In panel 3, for instance ($t = 2$ s), the airplane is hovering in the air although there is no building at the timespot and is thus regarded as a false positive. Panel 5 ($t = 6$ s) is a successful trial. Panel 10 ($t = 10$ s) is a miss, as the airplane is crashing into the Empire State Building).

This forms a multi-dimensional cluster in the channel space that records the activity in each of the channels during the presentation of each image. Future population vector samples are decoded by a distance function using the “Mahalanobis distance”²⁰. The sample is then labeled as belonging to one of the channels, or to the baseline activity (which can be looked at as another channel), and a decision is made according to the operation defined for that channel (“Go up”, “Go down”, for instance). Finally, the experiment computer runs the computer game. The experiment computer continuously sends independent triggers to the acquisition system indicating times of relevant events in the game and synchronizing the signal for future reference.

²⁰ Mahalanobis distance — a distance measure which takes into account the correlations of the data set and is scale-invariant, thus is not dependent on the shape or level of variance in each cluster. In our data it is useful because clusters tend to sometimes be very variable across the different dimensions.

Results

In order to test the ability of patients to control the airplane and avoiding the obstacles we measured two variables: a) the ability of the patient to indeed increase the firing rate in the required time (even if the patient was unable to increase the airplane above the required threshold height of the obstacle, but was able to increase the firing rate above the baseline at the time the building was approaching, it would indicate a level of voluntary control of the neuron, which is remarkable given that no evidence of direct access to the sensory system has been given before). This is regarded the “Permissive” criteria. b) the ability to increase the firing rate above the obstacle in the indicated time. This is regarded the “Restrictive” criteria.

In order to test the two variables we constructed 1000 surrogate random obstacle courses with the same amount of obstacles and the same heights for each, and tested how many times the airplane crossed the obstacles in these courses (restrictive), and how many times there was an increase in the firing rate when these obstacles were faced (permissive). For each obstacle course we were able to generate 4 numbers quantifying the success rate of the player in controlling the plane. The True Positives were the amount of 100 ms bins where the airplane was above the ground (above baseline firing rate of the neuron) when an obstacle was encountered. The True Negatives were cases when the airplane was on the ground in the absence of an obstacle (meaning – the patient didn't need to increase the airplane and indeed he did not), a False Positive is an increase of the airplane in the absence of an obstacle (notice that typically we would have a few 100 ms bins of false positive towards each obstacle approach as the patients commonly increased the airplane not just at the moment the obstacle was approaching, but a few hundred milliseconds earlier) – our measure in that sense is very

conservative, as it assumes the patient is rewarded only for not doing anything until the very same moment the obstacle is to be encountered (this is especially conservative since the patients were not asked to do that, but were rather told that they should avoid the obstacle in any way feasible to them). Finally, a False Negative would be a trial where the patient was either unable to increase the airplane when the obstacle was nearing (permissive), or wasn't able to achieve the required height (restrictive). For the illustration used in Figure 140 (patient 1) we see that while there were 28 obstacles in the course of the 180 seconds, the patient was able to avoid colliding with 20 of them (restrictive criteria), while he was able to increase his firing rate for 7 more – even though he could not reach a sufficient height to avoid them – getting a 27 out of 28 True Positives in the permissive criteria. The True Negatives rate for both criteria is 91 (out of 180 bins), and the False Positive for both is 61. The miss (False Negative) for the restrictive criteria is 8, and for the permissive is 1. These numbers were calculated for each of the 1000 surrogate courses and performance was measured as the fraction of Hits and True Negatives of the entire set of values: $Performance = \frac{TP+TN}{N}$. As such, if the patient correctly avoided all obstacles and didn't increase the airplane at all in the absence of an obstacle the performance would be 100%.

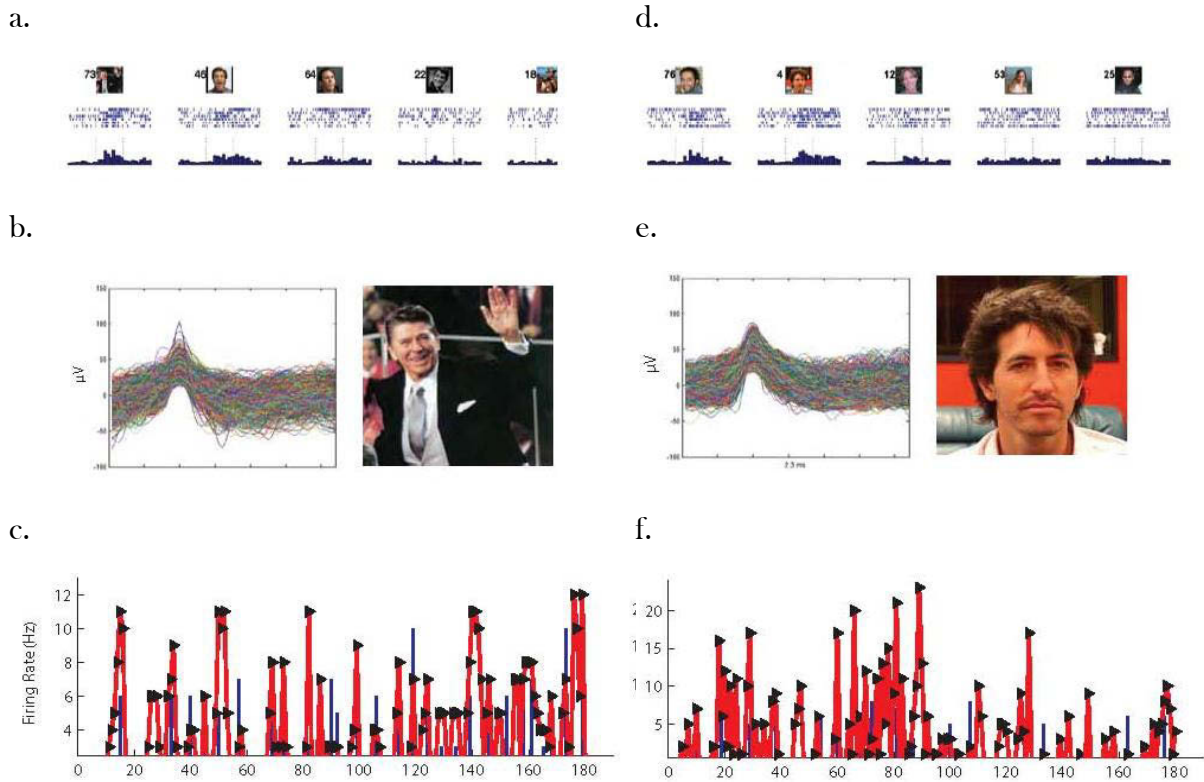


Figure 140 – Illustration of the game played and the single neuron eliciting the activity

a. Top 5 responses of channel 63 in the Left Anterior Hippocampus in patient 1. Best response is at the left. 6 trials' spikes are presented from top to bottom (first trial is up), where a dotted line on the left marks the onset of the image, and a dotted line on the right the offset. Image was presented on the screen for 1 s. Results are not clustered and as such include noise picked up by the electrode. Clustering the results yields cleaner and more significant results, however throughout the entire chapter we present unclustered data, as this was the feedback given to the patient in the course of the experiment. Best response is for the image of Ronald Reagan. See work of (Quiroga et al., 2005) for a detailed discussion of the screening experiment. Given the strong clear response to the picture in the screening, and a repeated follow-up experiment during the course of the day, we went to the patient's room later (9 hours

between the time the first screening started and the time the feedback trial began), and tested the ability of the patient to control the height of the airplane by thinking of Ronald Reagan. Patient had tried various types of feedback prior to the game and already gained some level of understanding with respect to the ability to control things using his thinking of Ronald Reagan. Overall the patient tried various types of feedback (acoustic – listening to the spiking activity in fashion similar to that of Hubel and Wiesel (“clicks” corresponding to discharge), visual – bar moving up and down, and such, for about 20 minutes). This was the first time the patient encountered the notion of feedback received directly from the brain, though, and the game was new to him. Overall the patient started playing the game no more than 30 minutes after he first learned of the notion of the “closed loop”.

b. Plot of 16037 spikes in the channel used for the feedback in the entire course of the 30 minutes of the experiment. Only very few of these spikes happen during the course of the game. This is to illustrate the level of noisiness in the environment we used. **Right panel.** The “concept” the patient used to control the airplane. Patient was playing with various notions and concepts that elicit responses based on his perception of the image. He tried thinking of other presidents, other republicans, focusing on different features in the image, repeating Reagan's name within his own mind, etc. The feedback session was completely unguided except for the fact that the patient was told that whatever drives this neuron has to do with **Ronald Reagan**.

c. The entire 3 minute games played by the patient. X shows the course of the game. At any given moment the patient sees a sliding window of a portion of the x axis (10 s), moving right every 100 ms. The y axis corresponds to the firing rate per second for the neuron, where the baseline activity is 2 Hz, and the maximal firing rate is 12 Hz. Black triangles mark the height

of the airplane at any given second. Black straight bars correspond to the obstacle building, with their height measured in Hz needed to avoid colliding with them. In this case, patient is able to avoid the first two building, but crashes into the 3rd. Overall, patient avoids colliding with 19 out of 28 buildings.

d. Neuronal response during the screening for patient 399. The patient shows invariant response for a person he was familiar with (response 1, 3, and 4), or for his written name (for elaboration of the method, see (Quiroga et al., 2005)).

e. The spikes shape for the course of the experiment. **Right panel.** the stimulus used for the game.

f. The game train for 180 seconds. For the plots we show the results binned in 1 s although they correspond to changes that occur every 100 ms. For the firing rate within each 1000 ms bin we use the sum of the firing rate within the 10 bins prior to it.

For each surrogate we calculate the performance and we test the null hypothesis that the distribution of the performances within the surrogates comes from the same median as that of the real game (Figure 141). In order to validate our significance measure we used the Wilcoxon test to perform a two-sided rank sum test of the hypothesis that the performance of the patient is the median of the performance of the surrogates. For both criteria and both patients, the null hypothesis were rejected at $p < 0.001$, suggesting that in both games the patients had control of the airplane based on the activity of the single neuron in their brains. This is true even when the task is to completely avoid running into the obstacles and not just to show any control of the airplane (restrictive).

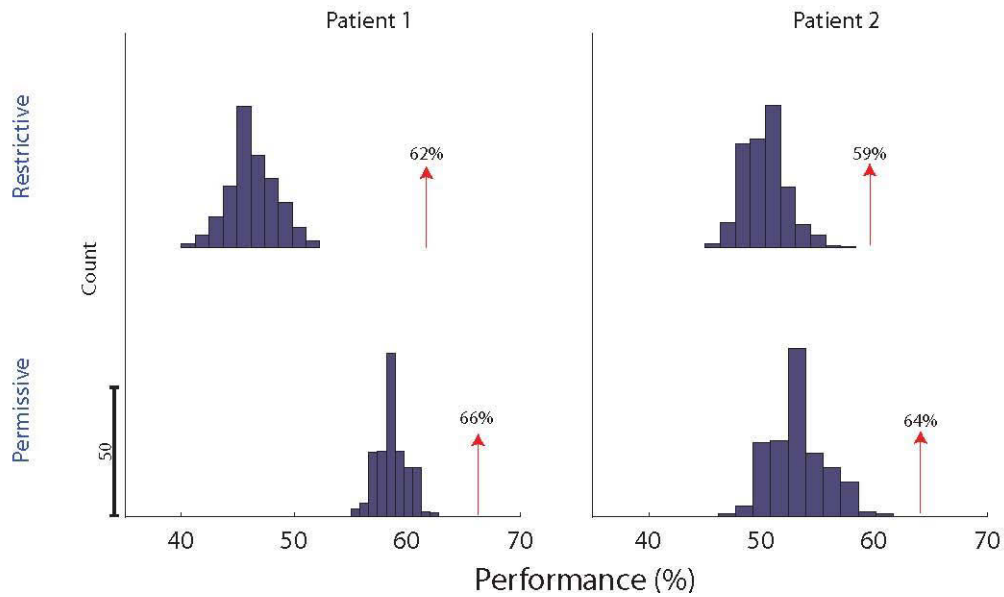


Figure 141 - Distribution presentation of the performance for the 2 patients

In 2 criteria (restrictive and permissive)

Left panel. performance measure for patient 394. Left distribution represents the distribution of 1000 surrogate obstacle courses created randomly from shuffling the real obstacle course, and testing the performance of the patients' path on them. The boxes have lines at the lower-quartile, median, and upper-quartile values. The whiskers are lines extending from each end of the boxes to show the extent of the rest of the data. Outliers are data with values beyond the ends of the whiskers. Star marks the performance of the real performance of the patient. Clearly the patient does significantly better than all 1000 randomly shuffled screens. The performance increase is 15%. The right plot corresponds to the same test with a more permissive criteria where the patient's maneuvers are regarded as "Hits" even if he was not able to avoid the collision – he is rewarded for just being able to increase the firing rate with the correct timing, and to not increase it when he should not. This was added since we could sometimes tell by looking at the screen that, while the patient is doing his best, he just could

not overcome the height of the extremely high bars (at some points the bars required a firing of the neuron which was 8 s.t.d. above the baseline). This method yields higher performance from the patient, but also from the surrogates, and while the patient still outperforms the surrogate, he does that by a smaller margin now. Clearly both cases are indicative of significant performance on the patients' behalf.

Measuring the performance change we see that the performance increase for patient 1 in the restrictive criteria is 16%, which is higher than the increase in performance for the permissive one (7%). Although the performance is obviously higher in the permissive evaluation, the difference between the performance of the patient and the surrogate is lower. For patient 2 the effect is enhanced. The difference in performance between the restrictive measure and the median of the surrogate is 9%, but, it is in fact 10% different for the permissive criteria.

In order to make sure that the dramatic demonstration of control of a single neuron by the patient is not a result of an overall change in the firing rate of the entire area of neighboring neurons, we performed the same test for performance in the task for each of the other 63 channels recorded. While the p value for comparison of the real data and the surrogates for the channels used in both patients were always lower than 0.001, the other 63 neurons in both patients were on average at 0.56 (patient 1) and 0.77 (patient 2), suggesting that indeed no other neuron recorded would show such high performance in the game but the actual neuron controlled directly by the patient (see Figure 142).

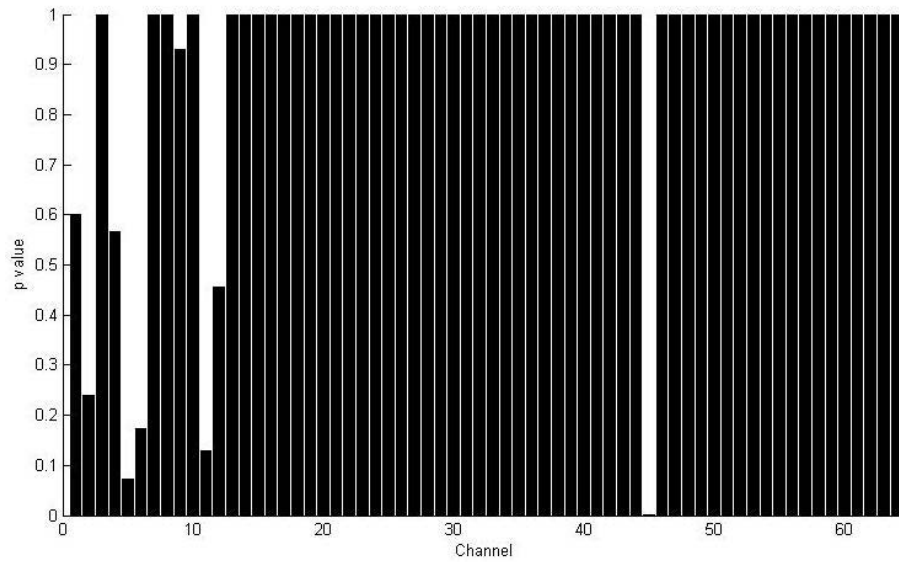


Figure 142 - Performance of neighboring neurons in the game

The p value for the comparison of each single neuron out of 64 to 1000 randomly created obstacle courses. Neuron 45 is the one actually controlled by the patient, yielding a very low p value, while the other 63 neurons show p values ranging from 0.08 to 1.

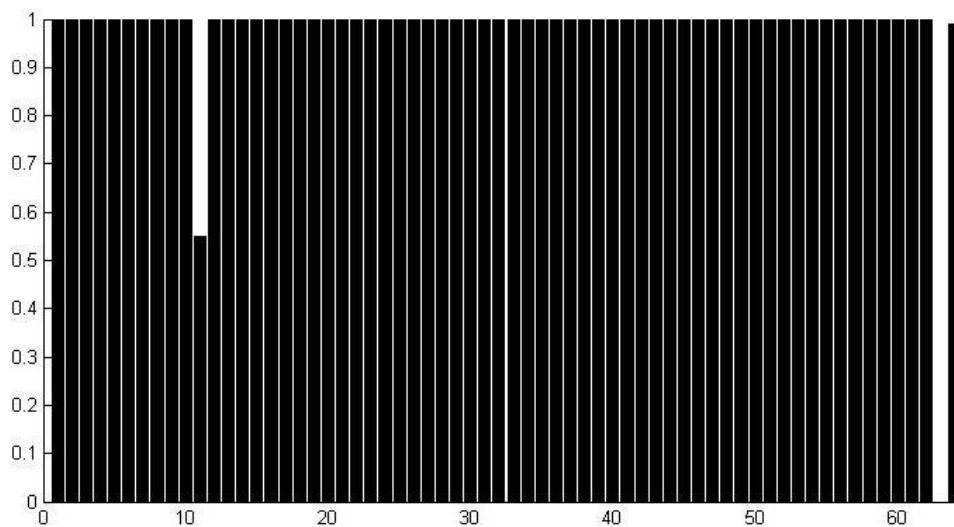


Figure 143 - Performance of neighboring neurons in the game

The p value for the comparison of a single neuron out of 64 to 1000 randomly created obstacle courses. Neuron 63 is the one actually controlled by the patient, yielding a very low p value, while the other 63 neurons show p values ranging from 0.55 to 1.

Finally, given the obvious significant effect of neuronal control for the computer game, we compared the amount of time taken to reach this high performance to that reported in comparable studies. While in (Hochberg et al., 2006) —regarded as one of the best examples of BMI with invasive recording method to this date with a human — the time to reach a level of correlation between brain-controlled cursor and a large ensemble of neurons was a few months (and the maximal correlation was 0.74 for the y axis ($r_2 = 0.45 \pm 0.15$) and 0.6 for the x axis ($r_2 = 0.56 \pm 0.18$), with average-to-maximal r^2 of 0.67), we reach a performance of 0.64 (patient 2, permissive criteria, as an example — this holds for both patients) after only 25 minutes of experiment with a single neuron. Hence, we show that the patients were able to reach high performance in comparable task within times that were much shorter (at most 40 minutes). In both cases, the training was shorter than the ones reported in previous studies with primates, or in the single reported study with a human tetraplegic subject.

Discussion

First, we demonstrated a real-time system with internal decoding algorithm that allows for visualization (or acoustic feedback) for the spikes recorded from human (and potentially animal, or simulation) brains. The system includes a decoder that allows for working with larger amounts of acquired channels and can make immediate decisions on decoded action. The system was tested with two patients in an engaging computer-game-playing task, and was shown to reach a high level of reliability, both in terms of data accuracy and speed. The system is scalable to larger sample sizes and as such is ideal for both decoding of sparse neurons as the ones presented in this study, or for larger ensembles as used commonly for motor-controlled BMIs.

Second, we show striking results with regards to the ability of patients to reach a high level of control of a single neuron in the sensory system that is not commonly directly accessed voluntarily. After a short training that lasted less than 30 minutes patients were able to reach significant accuracy above chance (average of 15% higher in a restrictive criteria for performance measuring). We show an engaging task performed by the patient that made them try to maximize their performance almost independent of their understanding of its effect on the neuronal code. Patients were able to alter the firing rate of a single neuron in the hippocampus that was chosen based on its selectivity for a single image out of hundred presented in a different study with just the power of their willful imagery done internally within their mind. Analysis of neighboring neurons indeed shows that the ability to control the computer game by a single neuron is due to a very accurate control of that neuron and not a general increase in activity.

The short time patients need to reach the high level of accuracy in control shows that performing a BMI task, based on imagery with a single neuron, can yield a very high performance rate based on volitional control of a single neuron. An increase in the size of the population vector, or more training time, should yield an even higher performance. This suggests that while our neurons are not ideal candidates for motor control, as they are concept neurons of things relevant to the patients' recent memory and are not closely tied to any aspect of movement, they can yield a very high decoding rate, due to their sparse coding.

It did not escape our attention that such conscious control of a neuron provides evidence for the ability to indirectly access areas of the brain that are not commonly accessed and can give rise to many interesting data that have to do purely with the effect of such task on the neuronal level (be it on the environment surrounding the neuron that might inhibit its activity to facilitate the task, be it in plasticity of the neurons that suggest evidence of learning of the system with respect to the task, or be it with general change of the network wiring in order to allow access to neurons that are not commonly triggered voluntarily). Such studies can benefit not only paraplegic patients, or people interested in brain-machine interfaces, but can also provide evidence of the ability of subjects to get some level of control of internal networks in the brain that can be influenced voluntarily. We can imagine a patient reducing his heart-rate using such "independent" or indirect feedback, or even affecting his pain level, if the sparse neuron that would be controlling the airplane happens to be coming from the "pain" system. This should be studied more and has very many potential applications to medical and behavioral research.

In conclusion, we presented here a study that is the first to show results of a BMI using single neurons in the Medial Temporal Lobe which are not directly part of the motor system for control of a motor task, and present both a technical innovation as well as a striking advance in the field of BMIs. This can be of use both for scientific study of the way this activity affects the performance of the neurons and the ability to access areas in the brain that are not commonly directly accessed, and as a method for using sparse neurons for decoding of free choices of patients and of concepts imagined by the patient in real-time.



Future

*Predictions are very hard;
Especially about the future*

Niels Bohr

Kay Kurzwill, inventor of the Kurzweil keyboard, and - since then - a self-declared “futurist” used to write a book every decade where he tried to predict the future. I will try to do it in a safer way, by suggesting a list of questions that would be interesting to follow-up on and answer using my data. These, I hope, will serve as a good guideline for a follow-up on my work. Similar to the famous kids game: “21 questions”, I list 21 interesting simple and immediate questions one can answer purely by using the available data.

21 questions

1. Evaluate the performance of subjects in the search task.
2. Evaluate the rating subjects give to the images in Block 5 which contain IAPS images (emotional). Take the outliers and see if they look at different things in the image.
3. Quantify the intra-subject correlation, when:
 - a. Viewing the SAME image within a block.
 - b. Viewing the SAME image across blocks, with different tasks (search / free viewing).
 - c. Differences with viewing time (first, second, third...)
4. Quantify the inter-subject correlation for the same (3a-3c).
5. Measure differences in viewing by gender of the subject / the person in the image.
6. Quantify differences in viewing of the SAME image based on color (compare Black and White images to color images - block 3 and block 1).

7. Compare the viewing patterns for all images of the **SAME** frame, with different locations of the face. This can shed light on the changes in saliency, as the saliency map for the same image would change for each image, for the same frameset.
8. Based on similar study showing that the fixation prior to the one landing on the target in a search task is also informative, calculate the location of the fixation **PRIOR** to the one landing on faces/text in the dataset.
9. Extend the fixations **ROI** in the images to features of the face (eyes, mouth, etc.) and identify the locations that trigger the most viewing.
10. Repeat (9) across subject groups (AgCC, Autism, etc).
11. Compare the **IAPS** outliers for Autism / AgCC subjects to their viewing patterns (i.e., if an autism subject rates a negative image as positive, does he look at different locations in the image).
12. Look (especially for the autism group) at the locations they fixate on **AFTER** they already looked at the faces. While controls 'stop' at the face, it seems that autism subjects would continue to scan the image, even in the search tasks, as they are less inclined to spend time on the face voluntarily.
13. In the search task block, in the trials where the subjects were mistaken (did not locate the face), quantify the proportions of their viewing the face (i.e., when they fail to say there was a face there, is it because they didn't see it - mistake, or despite seeing it - consciousness mistake).
14. Contrary to (13). When the subject is correct in identifying a face, is he always in fact fixating on it, or are these peripheral / gist identifications.

15. Create 'heat maps' for the emotional saliency in the IAPS images (block 5), and compare (ROC) those to the viewing patterns of autism subjects on the same images.
16. Same as (15) for AgCC and SM.
17. Compare the top-down / bottom-up saliency in the images cognitively, by having subjects mouse-click on the location in the image that they consider the most salient, and compare that to the saliency model and first fixations results for each group.
18. Quantify faces viewing in Blocks 5 and 7 (anomalies).
19. Quantify the features within a face (eyes, mouth, etc.) viewing within blocks 5 and 7 (anomalies).
20. Compare the anomalies images' saliency-fixations (ROC) to that of the same image without the anomalous changes.
21. Compare the rating of SM for the anomalous images, IAPS images and the regular images, and identify the viewing of anomalies altogether by her (as she missed multiple anomalies and rated them regularly).

Predictions

Finally, as I feel good scientists don't just end up studying the existing world, but also should make predictions about the future, and suggest investigations that need to be done in order to test those, I am willing to list my thoughts on the research and the future of my study that needs to be done in the coming 10 years.

1. I think that in the coming 10 years neuroscientists will eliminate more and more from the field currently known as “psychology” and will provide explanations to phenomenon which currently seem to be explained and addressed only on the ‘therapists couch’. For example, while nowadays the motives and inner explanations of a person’s behavior is typically explained by discussion of his actions, I feel that by simple tasks similar to the ones I demonstrated in my study a person would be diagnosed as part of a subgroup (like autism) of behavior stereotypes and accordingly will be offered treatment.
2. I believe that autism syndromes will be determined not by an elusive interaction with a therapist but by a clear diagnostic test similar to the ones I demonstrated in this work (subject would view a set of images, rate them, etc., and based on his score – proportion of face viewing / ability to decode images from his scanpath / etc. – he would be diagnosed as autistic.
3. I think that the similarities between subjects’ scanpaths will be fully explained as long as they are converging. That is, as long as the variability across subjects is small (see the ‘ideal-ROC’ mentioned in chapter 15) when viewing the same image, scientists will be able to fully explain those scanpaths.

4. Studies of the underlying mechanisms that govern our retinal movement will allow us to explain the earlier saccades (up to 60ms after image onset) to objects in the images purely based on features of the images.
5. The competitive mechanisms underlying human attention will be targeted as specific brain regions which are the main facilitators of those will be identified. The mechanisms - mainly an interplay between the prefrontal cortex, the frontal eye-field, the amygdala, and the superior colliculus - will explain the majority of the early attention allocation to subjects,
6. The field of eye-movement predictions will move out of the eye-tracking psychophysics room entirely and will provide explanation to the fixations and attention allocation of subjects in moving environment, with head-mounted light eye-tracking devices. No more will studies focus on grating and moving dots in a confined room, and natural scenes will be the majority of stimuli used.
7. The competition for attention will be explained to the point that marketing and advertising will focus on generating stimuli that will surely capture the attention of the observer for a quantifiable time (i.e., one would know that this particular message will activate this region in the brain for this amount of time - guaranteed).
8. Neuroscientists and marketing people will join hands in generating messages that are focused on emotions, brain regions and cognitive tasks of the observers.
9. Interplay of augmented-reality and vision will create new medium for stimuli viewing. Our focus of attention will thus change and we will be able to use better environmental cues to capture observers' attention.

10. Eyes-based devices will take a leading role in computation and instrumental uses. That is, many tasks that are currently done using our hands (windows focus changing, writing, driving, etc.) will be done purely based on our “where we look”, and “what we look at”.
11. Eye-tracking devices will be installed in our computer screens, cars, televisions, etc. and will have active role in assisting us rapidly activate devices.
12. Computer games will put this information to use, and gaming consoles will have games which target our attention shifts as part of the game.
13. While consciousness per-se will not be explained in the coming decade, and insight into the ability to alter the competition within a person’s brain for the ‘conscious thought’ will be explain. That is, we would be able to tell in advance out of multiple stimuli exactly which one will be perceived, at what latency, for how long, etc. We would therefore be able to study the early stages in life at which stimuli penetrate our brains (even with newborns) and understand the relation between what we learn and what we are exposed to.
14. Competition between brain regions for attention will be explain in altered states of consciousness (i.e., sleep, coma, etc.), but not in awake behaving humans.
15. New biological methods which become popular will nearly eliminate the usage of prior methods such as EEG, fMRI, etc., and will offer a better method to not only observe the competition for attention in the brain, but actually totally control it by activating and deactivating the mechanisms that interplay in order to determine which of multiple stimuli becomes the one we are conscious of.

That's it. Fifteen predictions for the study of attention, competition and consciousness between the years 2009 and 2019. I promise to report to the readership in 2020 on the level of accuracy in my predictions.



Conclusion

*There are two things you should
get in life: a wife and a Ph.D.*

*And between those two — the
Ph.D... no one can take it from you.*

Pietro Perona

When I started my research at Caltech, Christof gave me a few guidelines as to how my Ph.D. should be pursued. His main guideline was the following: “You should do such a work that – when you finish your dissertation – everyone in your field of study, as big or small as it may be – would know of your work.”

In a way this was an easy task for me. The initial field of work I chose was single-cell recording with humans. The amount of people involved in such a research can be counted on one hand, and they were all partially or fully involved in my research at one of its steps. That said, the vision part, and the general aspect of its relation to attention were not as trivial.

When I first started writing this conclusion I asked myself if I could accomplish the task or not. As I was debating within myself about this question I received an email. This email was from

From: Anonymous
Sent: Thursday, May 08, 2008 05:57
To: moran@klab.caltech.edu
Subject: paper

Hi There

I came across a PDF of your manuscript entitled 'Predicting human gaze using low-level saliency combined with face detection' whilst reviewing the literature about face perception. I had the chance to read the manuscript and I think it is REALLY interesting and certainly relevant to the research I have been doing. I was wondering if you could tell me more about your work as I am hoping to cross-reference it in a review paper I am currently preparing. I have also forwarded your manuscript to members of the Anonymous Face Perception lab where I was previously based as I have a number of collaborators there as well that would benefit from this work.

Best wishes

Anonymous
 Lecturer
 School of Psychology

an unknown professor in England, and it reads as follows:

Now I am not sure what this proves, but I do feel it's a good hint that this work is going in the right direction. Luckily, this was a first in a chain of various emails with the same attitude from people who I feel are expert in the field. A small confirmation that the little dent this work posed in the field is acknowledged by our peers. As Christof requested.

Summary

This thesis, as does my work, follows two paths, trying to answer one question: “How does the brain choose which of a vast amount of information that it inputs from the environment to become conscious of?”

Where in the brain is this unique bottleneck that decides that *this* sound will not reach our consciousness — that is, we will not be aware of it at all — while *that* sound will become the percept we will be aware of?

In order to tackle this question we used various methods to tap into the mechanisms in the brain that process the decision and used various techniques to “push the brain to some limits” where we would see this process happen under a controlled environment.

The two paths we used were tests with patients undergoing brain surgery, using direct recordings from neurons in their brains, and experiments consisting of presenting visual stimuli to subjects whose eyes were tracked to monitor the selection of information they choose to attend to in the images.

We started with a simple psychophysics experiments where we demonstrated that people tend to be consistent with themselves and with others in judging how conspicuous one image is out of set of hundreds of images. We noticed that people are not only consistent with their own responses, even over the course of years, but are also consistent with the responses of other people they have never met. This suggested an internal mechanism involved in our perception and decision about visual stimuli that kicks in especially when the exposure is short and controlled.

We used this information to construct an experiment where we showed people sequences of images (where we knew they would be consistent in their ranking and viewing of these) containing faces. We used faces as an attractor, as well as an important ecological cue to see if the fixations of subjects viewing those images would be as consistent with themselves and with others as their ratings of the images were.

We indeed noticed that subjects looking at images with faces are incredibly similar in their viewing patterns of the images, and are also incredibly predictable. They were in fact so predictable that we were able to pool together subjects and use their scanpath to pick the image they were looking at correctly out of hundreds, just based on their scanpaths.

In order to understand the level of attraction that faces exhibit in governing the attention when viewing scenes with people in them we conducted control experiments showing phones or text as other cues (text was a positive control in the images as another very semantic and important cue that we are exposed to continuously in our lives, but most likely are not born with the desire to look at; while phones are clearly not as important a cue in our lives). We demonstrated the uniqueness of faces in drawing the attention in nearly every scene where they

exist suggesting that there are some entities that penetrate this bottleneck of attention nearly each and every time they are in the scene. This argued towards some entities that always win the competition towards becoming the percept we will notice, or the thing our attention is drawn to when they are visible.

This, in turn, suggests some weights of the connections in our brain leading to this elusive conscious percept that can be tweaked using accurate stimulus.

In order to confirm this claim we went a step further and designed an experiment where we tried to have subjects control the percept they look at (which is a measure of the attention, as we claim one looks at the thing he attends to) by telling them in advance that images will or will not contain some elements that typically draw their attention, and have them make an effort to inhibit the need to saccade to these. In the “avoid” experiment we told subjects to try and not look at faces in images, and later not look at text or phones, showing that they are far better in avoiding phones for instance than they are faces. This, indeed suggests that the ability to avoid looking at faces is something that even a conscious will to override is not fully able to do. This more strongly supports our claim that attention is in many ways a process that happens independent of our control, although we argue that one is able to alter the weighting of the stimulus, and thus in fact channel the information flow in the brain into one or another direction.

Since we found the results of the faces attention attraction so striking we went one step further into prediction — which is the ultimate proof of a theory. We embedded a face detection algorithm in the saliency model to predict where people will look in the images they see. We compared the results of a model predicting the saccades and fixations of subjects, and were

able to reach a high level of prediction — for some subjects reaching over 90% correct predictions, while the average was at 80% — for images that contained faces. This prediction level is far beyond the prior prediction level for natural scene images, which was around 53%. This ability to accurately predict the fixation (we should note that this holds for images with faces, and mainly refers to the first fixations — when subjects are given enough time they start diverging and become impossible to predict) shows that the brain mechanisms that govern our attention in the early stages for certain scenes are constant between various people and therefore could be looked at as an innate mechanism for attention allocation in the brain.

As a further control for this claim we tested subjects with various brain disorders that show no attention allocation or interest in social scenes including faces. These were subjects with autism or agenesis of the corpus callosum. Our hypothesis was that attention allocation in the brain is indeed innate, and since these subjects do not share this mechanism, they will show differences in their viewing patterns for the given set of images used for the eye-tracking experiment. Mainly, their level of attention allocation would be smaller, and they would find other visual cues that are more conspicuous (feature-based cues, such as a glowing cellphone, or a very colorful Rubik's cube) more attractive.

Indeed, a study we conducted with subjects from these two groups showed a significant decrease in the level of attention allocated to face and social scenes in early fixations.

These confirmed our general hypothesis regarding the mediating of attention by bottom-up-based modules, and assisted us in clarifying the mechanisms behind such. However, in order to fully study the brain regions governing these mechanisms we used 3 different additional levels of testing.

The first test was done with a unique subject that has no amygdala. In Figure 93 we see that the amygdala acts as a center for the connections contributing to the attention allocated to faces and social entity. With the subject SM we demonstrated that while she lacks the amygdala, and the ability to identify many unique social scenes as important, she does exhibit a very similar pattern of viewing of the images, suggesting that the amygdala is not part of the system modulating the attention allocated to social scenes in the brain. More so, we showed that the level of personal relevance of the social scene affects the degree of attention allocated to it, by showing SM various images with herself and people she is familiar with rather than ones with unknown and nonpersonal objects.

The second test was done with two unique subjects that manifest a different brain disorder, although not as clearly and easily identifiable in the brain as the amygdala lesion. These are people with prosopagnosia – the inability to see and identify faces. While the results with these two subjects are not conclusive, as they show different behavior in their level of face blindness, we can clearly see the effect of their inability to identify faces on the performance of attention allocation. When timing was compared between these subjects and the controls we see that their attention to faces, if these are visited at all, is very different in nature. They scan the faces without looking at particular features in the face, and generally tend to look at faces less than controls or even autism subjects. Since our prosopagnosia subjects behave very differently (one almost doesn't look at faces at all, while the other does show some attention allocation to faces) it is hard to draw any clear conclusion about the attention allocated by these. More so, the lack of understanding of the brain regions that govern these two subjects' face blindness limits our ability to fully understand which brain mechanisms are the ones mitigating the flow of information regarding the faces or the social scenes elsewhere, or with

lower weighting – yielding a lower attention to these. The common belief, though, that faces are perceived in early visual system regions such as the FFA supports the bottom-up hypothesis regarding the early vision system's contribution to attention allocation and decisions governing the conscious percept.

Finally, in order to really understand the brain mechanisms behind the allocation of attention we looked directly at neurons of humans, and designed a situation wherein the brain regions which are suspected of reflecting the attention in the brain directly compete over the conscious percept. We had patients with epilepsy who were implanted with electrodes recording from single units in the human brain perform a task where they were shown two competing concepts and were told to focus their attention on only one. While the other one was visible and sometimes salient and noticeable we saw patients who were able to completely control the flow of information to independent regions (sometimes even on different hemispheres), suggesting that indeed the bottleneck of attention can be accessed voluntarily by a patient, and that the weighting can be directly changed based on inputs from within. More so, it suggests a real bottleneck region where decision about the information flow to consciousness happens – most likely in early visual system, before the information flows to either of the Medial Temporal Lobe regions. This indeed is a direct evidence for our hypothesis regarding the way attention is governing the progress of information in the brain up to the conscious percept.

Contribution and significance of this work

The contributions to the field of research of attention were in five realms:

At the cognitive level we demonstrated in a very quantitative measure that some high-level cues in the environment act as attention attractors in a bottom-up fashion. We showed that faces in natural scenes act as a major attractor, winning even against personal inhibition to avoid looking at those. While previous work used gratings, Gabor filters, or various very low-level features in experiments because of the assumption that high-level cues will increase the variability among subjects' saccades, we demonstrated the opposite effect. Using natural scene images with faces actually yields better predictions and lower variability across observers.

At the clinical level, we tested populations with disorders such as AgCC and autism, showing intriguing differences among the groups and controls. While we could not find a distinguishing clinical feature that allows us to tell if a subject is autistic purely based on our test, is it very likely that lower-functioning subjects with autism (especially ones who are not trained to look at faces as part of the treatment, which was most often the case with our autism) will actually manifest these distinguishing features, and thus a simple experiment like ours might be useful as an actual diagnostic tool.

At the neuronal level we demonstrated a first real-time system for detection of spikes that was used to conduct closed-loop experiments where human subjects in fact control their brains. This is among the first work in the field of that nature. The ability to reach a real-time control of any feedback, let alone competing feedback, is novel and remarkable. Patients in our

experiments were able to reach a performance higher than ever demonstrated in humans or other mammals, after an extremely short time and very few practice trials. Our subjects were able to control a single neuron in some cases perfectly to the point they could play a computer game using that neuron.

At the level of attention we demonstrated for the first time a task that allows for the creation of direct competition between brain regions and showed tendencies by one or other region to win based on patients' will.

Finally, we were able to tackle in this research one of the questions that baffled the study of attention and consciousness for many years concerning the proportion of effects on our percept due to feedback from the environment versus the effect that is due to internal brain mechanisms. Using a computer model fitting data from patients we could suggest a ratio of approximately 17% contribution to our conscious percepts due to internal mechanisms while the remaining 83% are suggested to be driven directly or indirectly by our environment. While this result is only suggestive, it should act as an important pathfinder and milestone in the search for the neural mechanisms governing attention allocation in the brain.

It seems that from the literature review I pointed out in the introductory chapters, especially the reviews about the competition model of attention, very many of the papers mention in their future questions and works the need for a direct experiment that will show competition for attention on a neuronal level with humans, as a direct experiment. This, to the best of my

knowledge is the first such experiment and as such should be a useful marker in this particular research.

Consciousness

While I consciously made an effort to refrain from addressing the concept of consciousness throughout my work, as this elusive term is very controversial in nature and might raise various concerns regarding its exact definition and spectrum, it is evident that this research — as I pointed out in the prelude chapter — began with and aims to answer this difficult problem: How is the conscious percept formed in the brain?

Attention in general is different than consciousness; however we clearly see that one is strongly tied with the other. There's a strong entanglement between the two, although very many experiments have shown that it is possible to manipulate the first without affecting the second. Our tapping into the world of consciousness is done mainly using the single-cell recording method. The ability to directly see the neuronal activity corresponding to percepts and concepts in the patients' brains is the ultimate manifestation of the conscious perception of those.

In a way, our brain control experiments show a unique and very unconventional reflection of our consciousness. That is, through a reversal of the usual cause and effect aspects of our perceptual flow. While we normally see the flow of information in the brain in the following order: signal from the environment is perceived by our senses, which in turn transfer it through various layers of evaluation to create a percept, which creates a thought or reflection on it,

which ends up in an action that might exhibit a functional motor movement that effects the environment (Kant, 1882); we here show a completely opposite order of “stream of consciousness” (Joyce, 1998; 2008 קרניוק). The patient first chooses (an external *active* decision, which has a corresponding internal brain process) the percept to think of, and then directly manipulates within his own brain the connectivity to select which of the relevant stimuli to perceive. This is the ultimate evidence for our ability to access areas in the brain that are typically looked at as ones that we do not have direct access to, like consciousness. At any given moment we perceive and do many things that we are not conscious of. Our digestive system operates continuously without a decision or any conscious control of ours. Our heart beats without our need to “tell it to” – it just does. So are also our memory and perception mechanisms, typically. However, our experiments showed that in fact – using a unique setup and little training – we can indeed generate activity in these areas using our own free will. From becoming the puppet of our own brain, we can – for a short while – become the puppeteer; the way we want to think of ourselves.

Final sentence

So how does the brain elicit a conscious percept from these enormous amounts of data coming from the environment? In this thesis I tried to show that at least one mysterious batch process which we call “attention” helps the brain select one thought from many and allows the conscious ‘I’ to deal and interact with this single percept at any given time. But the answer to how this one is translated to a percept, I think, can best be answered by the following four words:

I do not know.

Index

- AgCC, xi, xxviii, 91, 101, 102, 103, 105, 106, 199, 412, 417, 443, 444, 445, 446, 447, 448, 449, 574, 575, 590
- agenesis of the corpus callosum. *See* AgCC
- amygdala, lv, 107, 419, 637
- art, xxi, xxv, 202, 203, 232, 235, 624
- attention, xviii, xxviii, xxix, 59, 78, 93, 100, 101, 102, 104, 106, 108, 112, 113, 114, 116, 118, 121, 128, 137, 138, 139, 140, 144, 147, 148, 149, 151, 154, 156, 157, 161, 165, 169, 170, 186, 188, 190, 197, 208, 231, 235, 240, 241, 247, 248, 253, 255, 257, 263, 264, 265, 280, 286, 287, 289, 290, 291, 292, 295, 298, 300, 302, 319, 327, 344, 348, 353, 355, 360, 369, 374, 377, 380, 381, 382, 383, 388, 389, 391, 393, 394, 397, 402, 417, 421, 422, 427, 435, 443, 445, 448, 449, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 474, 477, 478, 497, 498, 505, 515, 524, 529, 532, 533, 534, 536, 537, 538, 539, 540, 541, 542, 546, 569, 577, 578, 579, 582, 584, 585, 586, 587, 588, 590, 591, 593, 607, 624, 625, 629, 630, 631, 633, 634, 635, 636, 640, 641, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 654, 655, 656, 657, 658
- bottom-up, xxviii, 79, 109, 139, 148, 149, 150, 151, 156, 158, 165, 170, 194, 208, 209, 230, 232, 245, 257, 263, 286, 291, 295, 297, 300, 319, 320, 321, 322, 332, 333, 339, 344, 345, 348, 353, 354, 355, 367, 369, 375, 382, 390, 395, 397, 398, 435, 454, 455, 456, 457, 459, 461, 462, 463, 464, 468, 469, 470, 471, 575, 587, 588, 590, 629, 633, 647
- top-down, 79, 139, 148, 149, 150, 194, 208, 235, 240, 269, 286, 290, 291, 295, 297, 316, 317, 318, 319, 321, 322, 332, 355, 398, 454, 455, 456, 457, 459, 462, 468, 469, 471, 541, 575, 633, 639
- autism, xi, xxviii, 78, 92, 93, 94, 95, 96, 98, 100, 106, 375, 380, 381, 382, 383, 384, 386, 387, 388, 389, 390, 392, 393, 394, 396, 397, 398, 399, 405, 408, 412, 417, 437, 438, 439, 441, 442, 443, 444, 445, 446, 448, 574, 575, 576, 586, 588, 590, 628, 629, 633, 638, 639, 644, 645, 647, 650, 651, 658
- Caltech, xi, xii, xiii, xiv, xvi, xviii, xix, xx, xxi, xxiii, xxiv, 62, 71, 72, 73, 74, 78, 91, 96, 107, 118, 161, 199, 202, 203, 211, 249, 266, 303, 380, 386, 419, 426, 547, 582, 597, 598, 623, 625
- Christof, iii, xiii, xxi, xxiv, xxv, 62, 72, 73, 74, 114, 156, 158, 380, 582, 583, 599, 623, 625
- competition, xxvii, xxviii, xxix, 78, 101, 102, 106, 109, 135, 139, 140, 141, 143, 145, 147, 148, 149, 150, 151, 156, 289, 300, 312, 313, 315, 319, 322, 328, 335, 417, 448, 474, 476, 484, 495, 500, 501, 502, 526, 578, 579, 585, 591, 629, 633
- consciousness, xv, xxiv, xxvii, xxix, 59, 67, 68, 72, 75, 123, 574, 579, 580, 583, 589, 591, 592, 642
- entorhinal cortex, 492, 494, 495, 509, 551
- epilepsy, xxviii, 78, 91, 109, 124, 125, 126, 127, 128, 197, 474, 479, 551, 588, 629, 637, 646, 650, 652, 653
- eye-tracking, xxvii, xxix, 78, 158, 182, 183, 184, 186, 187, 193, 237, 266, 271, 300, 303, 360, 363, 377, 385, 398, 407, 408, 411, 419, 423, 429, 431, 434, 437, 456, 457, 461, 462, 463, 468, 577, 586, 598

- faces, xv, xxvii, xxviii, 78, 100, 109, 114, 115, 119, 120, 121, 139, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 176, 179, 186, 187, 190, 235, 236, 237, 238, 239, 240, 241, 242, 243, 245, 247, 258, 264, 266, 267, 269, 270, 274, 275, 276, 277, 279, 280, 281, 282, 283, 284, 285, 286, 287, 289, 290, 291, 295, 300, 302, 306, 307, 309, 310, 314, 316, 317, 319, 320, 322, 323, 327, 333, 334, 340, 341, 342, 344, 347, 348, 349, 350, 351, 352, 353, 354, 355, 361, 363, 364, 367, 375, 380, 381, 382, 383, 384, 385, 386, 388, 389, 390, 391, 392, 393, 394, 395, 397, 398, 404, 407, 408, 411, 412, 413, 415, 416, 417, 419, 420, 422, 423, 424, 425, 426, 428, 429, 430, 432, 433, 435, 439, 443, 445, 446, 447, 448, 461, 574, 575, 584, 585, 586, 587, 590, 599, 600, 601, 606, 611, 612, 616, 624, 628, 629, 631, 639, 644, 645, 647, 653, 657, 658
- fading. *See* feedback
- feedback, xiii, 148, 384, 422, 434, 476, 486, 487, 491, 496, 500, 503, 504, 505, 506, 512, 515, 526, 529, 534, 537, 541, 542, 551, 552, 553, 561, 562, 568, 569, 591
- FFA, lv, 170, 588
- fusiform face area. *See* FFA
- game, xxi, xxix, 81, 470, 479, 505, 506, 524, 546, 549, 550, 551, 553, 554, 556, 557, 561, 562, 563, 565, 566, 568, 569, 573, 579, 591
- God, 66, 74, 525, 620
- hippocampus, 107, 131, 482, 483, 492, 495, 512, 514, 528, 551, 568
- memory, ix, xxii, 93, 94, 98, 104, 148, 188, 206, 209, 214, 216, 217, 224, 225, 227, 230, 237, 385, 393, 452, 456, 457, 526, 554, 569, 593, 612, 635, 638, 645, 648, 658
- parahippocampal cortex, 492, 494, 495, 512
- prosopagnosia, 78, 199, 411, 413, 414, 416, 417, 588, 628, 634
- saliency model, 79, 162, 186, 190, 194, 245, 291, 332, 333, 334, 338, 341, 344, 348, 354, 355, 359, 364, 382, 417, 461, 462, 498, 575, 585, 606

Appendix I. Statistics

- Number of times the word god is used in my thesis: 11
- Number of times the word basically is used in my thesis: 6
- Number of words in the thesis: 132,418
- Number of days spent writing this thesis: 367
- Number of days spent choosing the font and style for the thesis: 3
- Number of Mb the thesis Word file takes: 24.93Mb (26,210,597 bytes)
- Number of times the eye-tracker was calibrated during the course of my research: 881.
- Number of versions for this thesis: 23 (April 26, 2009)
- Number of papers written in the course of my thesis which were never submitted anywhere: 2
- Number of times I mistook the word “subject” for “patient” and vice versa in this thesis before manually correcting them all: 19
- Number of pages the thesis was before I learned that Caltech requires double spacing rather than 1.5 lines spacing: 421
- Number of days working on my thesis from the day I declared it completed: 90

- Number of Terabytes used to backup my single-cell data: 8.3 Tb
- Number of laptop computers used in the course of the research: 10
- Amount of money spent paying subjects for eye-tracking studies (estimated): \$1850
- Number of miles spent driving between Caltech and UCLA during the research (estimated): 8400 miles (20 patients x 7 days x 2 directions x 30 miles)
- Time it takes to print the thesis once the command “Print” was launched (at the Klab printer): 69 minutes
- Monetary value of the thesis (based on \$0.13 cents per page): \$79.56

Appendix II. Getting the Data

All the data for my analysis is available for whoever wishes to follow up on this work.

There are two databases created for this work:

1. **The Faces dataset.** <http://www.klab.caltech.edu/~moran/db/faces>

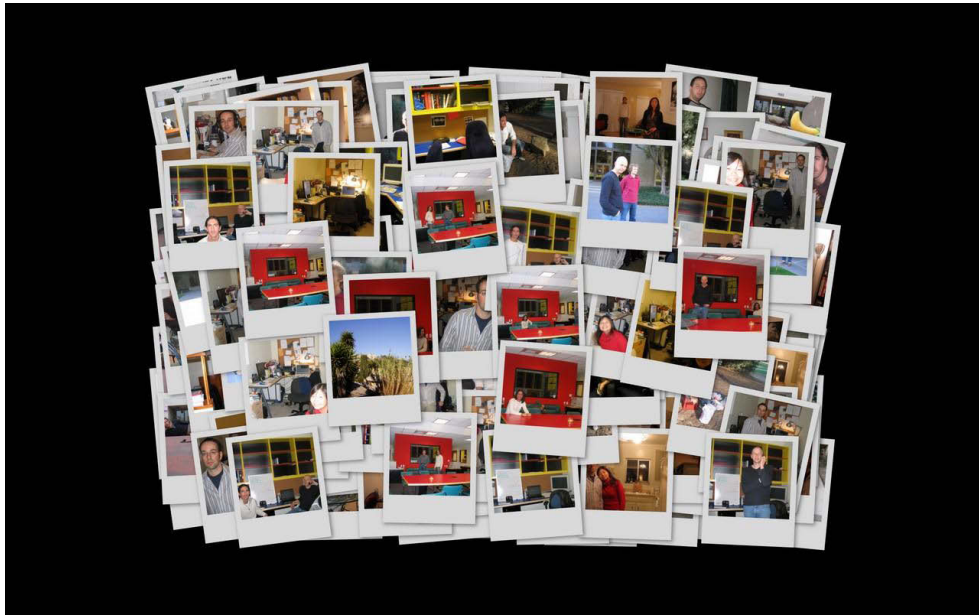
The faces database is a set of 1024x768 pixels images which contain frontal faces in various sizes, locations, skin colors, races, etc. Each frame includes one image with no faces for comparison.

The database is free for academic purposes. Please acknowledge the use of this database by citing this paper:

Cerf Moran, Harel Jonathan, Einhaeuser Wolfgang, Koch Christof, "**Predicting human gaze using low-level saliency combined with face detection**", *Advances in Neural Information Processing Systems (NIPS)*, 2007.

The database is broken to 4 categories:

1. Faces (230 images) - a set of images with faces.



2. Black and White (50 images) - a set of black and white images with faces.



3. Normalized exposures (29 images) - a set of images containing faces, with normalized exposures, for better comparison with saliency models that are based on luminosity and contrast measures.



4. Anomalous (66 images) - a set of images with odd variations of facial features and locations (e.g. eyes in abnormal positions, missing faces, faces shifted upside-down, etc).



2. **The Exposures dataset.** <http://www.klab.caltech.edu/~moran/db/exposure/>

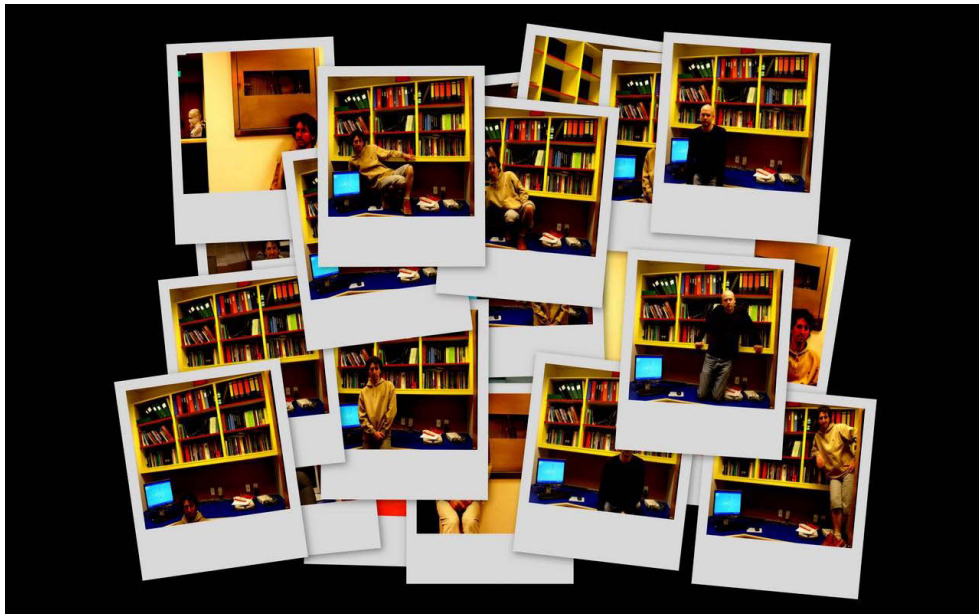
The exposures database is a set of 1024 x 768 pixels images in different controlled exposures. Such images can be used for better comparison with computer models that are based on pure image luminosity and contrast.

The database is broken to 4 categories:

1. Normalized images (29 images) - a set of images that were taken explicitly with controlled exposures so that no area in the image is too exposed.



2. 20% under exposures (29 images) - the same set of images, controlled such that all of them are exactly 20% under exposed.



3. 10% under exposures (29 images) - the same set of images, controlled such that all of them are exactly 10% under exposed.



4. 20% over exposures (29 images) - the same set of images, controlled such that all of them are exactly 20% over exposed.



5. 10% over exposures (29 images) - the same set of images, controlled such that all of them are exactly 10% over exposed.



3. The Fixations dataset. <http://www.fifadb.com>

The fixations of 8 subjects “free-viewing” the first block of 200 images (150 with faces) is made public and available for further studies and research at a dedicated website. Fixation In FAcEs DataBase – fifadb.

Fixations In FAcEs
DataBase

Introduction

Data

Images

Code

References

Contact

Updates

Faces

This website is dedicated to sharing data pertaining to observers viewing of faces in natural scenes. The data provided includes subjects' viewing of images containing faces, showing that faces are attracting the attention. Fixations are provided in Matlab format. Additionally, a Matlab code used for determining the saliency of locations in given images is provided. This code is used to predict the locations of subjects fixations in images.

For updates on changes/additions to our code/dataset please [register](#).

Featuring

Eye-tracking data used in various experiments
[net data >](#)

Code for analysis of bottom-up saliency with information about locations of faces
[net code >](#)

updated: August 17, 2009

The database holds the Matlab codes for running the saliency model on the given images, an extension of the code that enables the user to use the additional face detection implementation of the Viola and Jones algorithm in order to add the face channel to the saliency map.

Additionally, two Matlab files – one including the annotations of all the images, and one with the fixations of subjects viewing the images – are provided with examples and usage instruction to allow the scientific community to put the data to further use.

Appendix III. Interview with SM

MC. First of all tell me how you found the experiments this morning ?

SM. They are ok.

MC. Let's focus on the one you did with me. I want to know how you felt during the experiment ? What we are interested in in this experiment is comparing the way you look at things that you care about to things that you don't care about as much. We want to know what attracts your attention more. Do you think there is any difference in the way you look at things you care about vs. things you don't care about that much ?

SM. I think I am a little bit critical of me, so the images where I personally was in I cared less about. Several reasons: I... em... I feel like I look older than I really am. And part of the reason for that is my skin condition.

MC. So you say you look older than you really are? Do you think that you will look at yourself less because of that? Would that affect the things you care about in the images?

SM. Yes. Well... I am a woman. Women want to look beautiful forever. Well... not necessarily forever, but I want to look my age, and I want to not look above my age.

MC. So say you were looking at a picture. You think you would focus on aspects that reflect your age more for instance?

SM. Yes. I would look at my wrinkles... I want to look ■, not 65 or 70.

Entering my car

SM. This is a nice new car. I've never been in a convertible before.

MC. New experience, ah ?

SM. Yes. I like new experiences.

MC. OK. S. I have a couple of background personal questions to ask you. You don't have to answer, only if you're ok with that. OK ?

SM. I would answer everything. I am not embarrassed by anything.

MC. That's good. I just want you to feel ok not answering if you want. So tell me, S., how did you meet [REDACTED] ?

SM. I met him at a club, in [REDACTED]. I was 18.

MC. You went by yourself to a club ?

SM. Yes. I went there. Started shooting pool. He came out asked me if he can shoot pool with me. We started talking and...

MC. ... and you started dating right away ?

SM. Yes

MC. How long have you been dating him ?

SM. [REDACTED] years.

MC. Within which you had [REDACTED] babies. The [REDACTED] are from him ?

SM. Yes

MC. Good. Now I want to talk to you about your condition. I want you to tell me what do YOU think is the effect of your condition? What do you think having no amygdalae means?

SM. I think that maybe I think of things in a different way than other people?

MC. Explain to me what that means. Like... give me an example of something that you think you will think of differently than me?

SM. It's in how we judge things.

MC. What do we judge differently?

SM. I mean... it's in how we perceive people.

MC. Do you think it has to do with EVERY thing that we judge, or only people? For instance, we are now in a crowded place. Do you think we will have differences in the way we attend to thing in this place?

SM. No. It's only for particular things. But I truly don't know. It's a hard question. The thing about me, and I don't know if it's true for everybody else. Is that I like to challenge myself. If someone tells me that there's something I cannot do, I try to do it just to prove a point, so many things that I naturally am different in, I now do just because people say I cannot.

MC. Give me an example.

SM. My walks. I walk 25 miles a day. When I first started it, people didn't think I could do it. ■



SM. I would. I want to know what he says about me.

MC. So... I am going to wait a bit before I tell you everything about them, because we still want to test you under some conditions and not tell you what we study, but generally I want to tell you a little of what's written about you in those papers. I want to tell you what we study. We want to know, simply put, what the amygdala does. How it contributes to the way you look at things, and how you interpret emotional things ? Do you know what amygdala is?

SM. - It's my calcium thing.

M - Yes. Exactly. It has many functions, this part of the brain. And we care about the functions that it carries. Our experiments target all kinds of functions we think the amygdala is responsible for, and we ask what you – having no amygdala – will do in these. There are multiple answers: it could be that you perform exactly as the controls do, in which case we can say that not having an amygdala didn't affect the results. You might actually do better than us in many tasks, and we might say that having no amygdala improves your performance, or you might have trouble with tasks. This will suggest that the amygdala is helping us in these. Do you understand that?

SM. So why is showing faces relevant for that?

MC. I can't tell you everything now, but let me tell you a bit. One of the things we care about is how you perceive social things, so it's important for us to have you look at things that are social, like faces. Do you think you find faces and people interesting?

SM. Some of the things I saw I found pretty boring...

MC. Yes.... usually since we need a lot of data, we need to show you some things many times and it can become boring... I understand. Tell me S., what would YOU want to see? What images or things can I show you that you will find interesting to look at?

SM. I like scenery stuff. Mountains and such. I liked the picture with the large group of people that I knew. That was a good photo. [redacted] [laughing] I'm joking. [redacted]

MC. [redacted]

[redacted] Anyhow, going back to my questions - do you remember the experiments you did last time you were here?

SM. No... not really.

MC. How about the ones you did in [redacted]?

SM. I remember the MRI, and the emotions ones, but not much more.

MC. I am surprised. You have a very good memory, so it seems, from my memory experiment of yours.

SM. I just have difficulty remembering names and people sometimes.

MC. How about faces, do you remember faces?

SM. I remembered the faces of the people you showed me that I met before. People I liked.

MC. Do you have many friends? Say, in [redacted], do you have people you trust and confide in?

SM. I am pretty much alone.

MC. So most of the day. Who do you talk to?

SM. No one.

MC. Do you find it hard to make friends? Like... do you want to make friends but just find it hard to make them, or do you not want to make friends?

SM. I am pretty much a loner. I talk to people, and I often interact but I don't have that many friends...

MC. I am asking because I am curious to understand who you often talk to about your life. How about [REDACTED]. do you talk to [REDACTED] often ?

SM. Well... I have [REDACTED] that I would love to talk to more often.

MC. Do you find it hard to talk to people? For instance, talking to me now - is it a demanding task for you, or it is easy and trivial for you ?

SM. I find it easy. I enjoy meeting people. I enjoy talking to people. But being in a room with somebody it is hard for me to understand what they think?

MC. Why do you think that is?

SM. Because most of my life I put my trust in people and they always let me down.

MC. Let's talk about trust. We think that the part of the brain you miss - the amygdala - in a way assists you in correctly judging people. It helps you judge them better. And maybe if you lack the amygdala you are different in the ways by which you judge people. Maybe you are friendlier to people. What do you think?

SM. No. I don't think so. Let me ask you - when you first met me, how did you judge me ? and I don't mean as a doctor, or as a scientist. I mean as a person. Did you like me ?

MC. When I first met you, I knew I had clear borders with regards to you. I knew that there are some things that I cannot talk to you about. Question that have to do with your life or with science that I could not ask. I knew that I know a lot more about you than you know about me. I knew about your background and your history. I knew what you'll like and not like. So it was an unbalanced interaction from the very beginning. It makes our interaction very unique and makes the structure of it very unique. I know exactly what the limit of our relationship is, and as much as I am open and friendly with you I also feel that I am responsible for you. I know that you tend to be very friendly with strangers, and from the very beginning I knew that it was my responsibility to take care of you and make sure you are ok with the situation and to make sure no one takes advantage of you. For instance, yesterday when you were in the lab, it was very important for me to keep your condition secret so people in the lab won't judge you. I also - you might not be aware of that - but I also speak slower so you will understand me better. Usually I speak much faster. So there are all kinds of conditions that are clearly set when I interact with you. So you can't learn much from my interaction with you about normative relationship, and how people really judge you. So let's go back to you, though. In the interaction between us, do you feel comfortable?

SM. I wonder what is it that you can learn from me? I am curious what you learn? The thing about my condition, though, is that I am always very honest. If I felt uncomfortable with you - I would tell you.

MC. Would you?

SM. I would. I am not afraid to tell you. That goes for all my life. I am never afraid to say things. I have a problem lying. I am not very good at lying. I always tell the truth. That's the way I am. That's my experience, because of all I have been through.

MC. But... wait... there must be some people you don't like. What do you say to them?

SM. I would say: I don't like you. Stay away from me.

MC. Do you actually recall such a story? A story where you actually told someone to stay away from you?

SM. I do.

MC. Tell me about it. Tell me one such story.

SM. I was one time at a bar. And a guy kept coming up to me and getting a little too close. And I didn't find him attractive; I felt... how I should put it... I felt I could not trust him. So I went up to him and I simply told him: "stay away from me".

MC. So you don't have problem deciding what you want, and saying it.

SM. It's called standing up on your feet. Protecting yourself.

MC. This is interesting to me, because I thought that one of the things that you will have due to not having an amygdala is this lack of fear from people, and the tendency to, say, trust everyone. I thought it will be harder for you to keep people away than it is for me. I guess this is not true.

SM. Here's a challenge for you. Put me to the test. Take somebody that I don't know, and test my ability to tell if he's trustworthy, and then maybe you'll understand me a little bit more.

MC. Do you think that you are predictable?

SM. Depends... what do you mean?

MC. I'll be clearer. Do you think your reactions are predictable?

SM. I don't know.

MC. How about for yourself? Are you predictable to yourself?

SM. I think I am. I have routine, for instance. I like playing pool. That's what I do. I shoot pool. I know exactly how I will feel while shooting pool.

Food is served

MC. Ok. So I have another question. Are you scared easily?

SM. I am very hard to scare. I don't know what fear is.... Just yesterday, I was asked to identify the fearful faces... and I don't know what fear is. I could not identify it - the expression.

MC. So what did you mistake the fearful faces for?

SM. I could identify it - the expression. I don't know what fear is.

MC. What about the pictures of, say, the spider in today's experiment. In the pictures. did you find these frightening?

SM. No. not at all. I like spiders.... I am not scared of them.

MC. Are you telling me you are not afraid of anything? Do you mean you are very very strong? that you can OVERCOME the fear, or you really don't feel the fear at all ? Do you understand the difference?

SM. I am very strong.

MC. Can you think of something that made you afraid? Something you afraid of?

SM. In school, I used to beat the boys. I was never afraid of them....

MC. How about heights ? are those frightening to you ?

SM. No. not at all.

MC. You know... the feeling of fear has a purpose. It is there for a reason. It is there to help us avoid dangerous things. So although you can overcome the fear of heights for instance, it is there to help us avoid being in a high place because there's the chance that we will fall and die. So you think you can UNDERSTAND that feeling, the fear of heights?

SM. I am not afraid of heights.

MC. I understand that. But I am asking if you can understand why I, for instance, would be afraid of heights? Or why I would be afraid of spiders?

SM. I don't understand that.... I had too many things happen to me in my life. And - you must understand - if I wasn't as strong... if I didn't stand up on my feet, then me and [REDACTED] would have never survived.

MC. I remember your story from yesterday. You told me you stopped being afraid because you had to be strong. But I am trying to understand if it was a DECISION that you made because you had to survive, or if you really don't have the FEELING of fear. Like when you see a spider...

SM. ... I'd step on it.

MC. But do you tell yourself "I have to be strong", so I will overcome the fear and fight it, or do you not feel anything?

SM. I really don't feel the fear.

MC. Can you remember a time when you were scared of things? as a child maybe?

SM. No... I don't remember ever being scared.

MC. Can you make other people afraid? Say you were working in a haunted mansion in the circus, having to scare kids wandering around. Would you be able to do that?

SM. I probably could. I understand fear of others, I guess..

MC. Let me give you an example of what I mean. When I was younger, say 17 or 18, I used to be afraid of heights. Then I forced myself to go climb mountains and be in high places commonly. I was rock climbing often. I did it for so much time that by now I really am no longer afraid of heights. I overcame this fear. But the fear was genuinely there at first. I no longer feel it but I know what it is to be afraid of heights. Do you think that your lack of fear is more like this or just having absolutely no understanding of what's scary about being high?

SM. If I have to guess why people are afraid of heights, I would say that people do not want to get hurt.

MC. Right. So tell me - why do you think they are carrying this extra feeling of fear when it has to do with getting hurt? Do you yourself not feel that sometimes you can be hurt?

SM. I think that fear is a very bad thing. If you show fear then no one will ever leave you alone. I think that showing fear is not good. You know... folks will see it and you will die. But if you stay strong, and show that you're not going to be scared, then you'll be on your own.

MC. Your [REDACTED] Are you worried about [REDACTED] ?

SM. Yes. I am.

MC. What are you worried about?

SM. [REDACTED] ?

MC. No.

SM. I'll explain to you. I am worried about [REDACTED]
[REDACTED]

MC. See... there's something interesting there. You basically say that if someone held a gun to you - you won't be afraid. But if someone did that to [REDACTED] - then you are afraid.

SM. I am not afraid. I just don't want that [REDACTED] What you need to understand is that I am worried, but not afraid.

MC. What do you think is the difference between “being worried” about something and being “afraid” of something? Can you explain to me this difference?

SM. Afraid means being frightened. Being scared. And “worried” means not wanting something to happen. I have always been worried about things, but I am never afraid. If I could stand between [REDACTED] I would do that because I am not afraid and I am worried about him. I am not afraid about myself.

MC. So who is protecting you ?

SM. God.

MC. Only him ?

SM. Only him.

Appendix IV. Brain for Sale

One of the weirdest emails I got during my research (selected out of very many, as I was asked to be the lab correspondent for interacting with people who wanted to volunteer for the lab):

From: strange@email
To: moran@klab.caltech.edu
Subject: Brain for Sale

So I [was] browsing Google Videos and came across a lecture that Professor Koch gave a few years back on the topic of Consciousness. He mentioned that he was able to piggy-back data from epileptic patients who were undergoing "invasive brain surgery" but time was limited to the persons' stay and/or willingness to participate. Well, for a large sum of money (which we can discuss in further correspondence) I am willing to subject myself to this same surgery for as long as you need. You can put as many holes in my skull that you need just as long as it's still strong enough to take the occasional fall without shattering. Hell, you can custom make some of those tubes to have adapters at the end and leave the tubes in my head indefinitely (naturally, this will cost more). But then all we would have to do when you wanted more data is plug me into the machines and we'd be good to go. Just think of all you can discover with the data I'm offering. Nobel Prizes will be had.

If there are legal issues with this sort of thing, surely you know some trustworthy doctors who can diagnose me as epileptic. If we need to leave the country to make this happen, that works too. Just tell me where to sign. I eagerly await your response but until then, I remain, quite seriously yours,

Strange person

Appendix V. Patent

During the course of our work we also tried once or twice to see if our scientific results could actually yield some financial or industrial profit. That is, can they actually be used by people outside of the scientific community. One of these trials led to a suggested patent invention that we pursued for a year between my 4th and my 5th year at Caltech. Paxon Frady, Christof, and I tried to think of applications to the study corresponding to the usage of face detectors with a saliency map. Although eventually we decided not to pursue this work. I hereby list the main ideas that came about during our discussions of the idea, for an enthusiastic reader to consider picking it up and making into an actual device.

Ingredients

1. Saliency model
2. Rapid face detection algorithm
3. Algorithm for combining the conspicuity maps with accurate weight as to detect the most looked-at and attended-to areas in a given image

Suggested applications

Using the abovementioned ingredients, we were looking for ideas for applications that in fact would put to use the need to know where a person is, in real time, and use that information to locate the area at which the average observer would choose to fixate.

These are 3 suggested applications we had in mind:

1. **Supermarket merchant placement**

Organizing the shelves in a supermarket is a marketing and sales art. Supermarket organizers tap into the very inner shopping vices of all of us and turn our desires and cravings into money and profit. While organizing supermarkets could benefit from a saliency algorithm *per se* it does not need additional face/person identification. However, many supermarkets nowadays, on top of the shelves and products exhibited, have a few sales people standing in various locations in the supermarket offering samples. Now the placement of these people could in fact benefit from a saliency map with faces identification, suggesting the most noticeable locations at which these sales people should be put in order to be most visible and capture the attention of visiting customers.

2. **Automatically guided concerts camera**

Live-concert filming is a skill that's not easy to acquire. On top of a rapid technical skills and a good eye for visuals that will be relevant and intriguing to the audience, a director needs to identify in real-time all the interesting aspects of a scene that sometimes contain as many as 5 key people. The audience usually sees at any given moment, on a live screen, only one of the people, and various cameramen are in charge of moving, dollying, rotating, zooming, and controlling very many cameras on the stage. A single camera which will be able to identify the people in the scene, while choosing the most intriguing, conspicuous, and noticeable locations to film could prove to be of a remarkable help for a director — knowing that he can trust the moving cameras to always focus and center on the most interesting aspect of the scene which they are filming. An automatic guided concert camera, controlled by our algorithm may well be a useful tool for such.

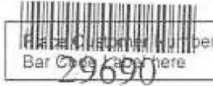
3. Casino surveillance


Casino cameras have 2 things they try to track at any given moment. People (as these are the ones suspected of cheating or misleading the casino in fraudulent acts), and generally conspicuous activity (governed by suspicious movement, abrupt change in intensity of the scene, etc.). These two combined seem like an ideal combination for a saliency-guided algorithm that uses people detectors. A camera that is based not only on the pure input from the scene, but actually picks relevant locations and scenes/people to isolate and focus on and suggest potential locations worthy of careful attention could well be a useful additional tool for casino security.

Out of these options, the one that generated most attention and interest from people we discussed the idea with was the automatic concert camera. To the best of my knowledge, to this date, it does not exist, and simple sampling of my friends working in the entertainment industry, this one can actually have a commercial potential.

While after consulting the Caltech attorney and Intellectual Property department we decided to not invest the money and time in pursuing the patent, I leave it here as part of my thesis for an enthusiastic entrepreneur to pick it up and make it viable. The main reason we decided to halt the process at some [rather advanced] stage was that we figured out that for such an idea to become a product someone must dedicate his full attention to it. Neither Christof nor I wanted to be that person, and the idea remained orphan. A courageous entrepreneur might find it useful to pick up where we left off. Figure 144 has the patent application.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET
 This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).
 Express Mail Label No. ED 273989263 US

INVENTOR(S)		
Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Christof Moran Paxon	Koch Cerf Frady	Pasadena, CA Pasadena, CA Pasadena, CA
<input type="checkbox"/> Additional inventors are being named on the _____ separately numbered sheets attached hereto		
TITLE OF THE INVENTION (280 characters max)		
Automatic Prediction of Human Gaze in Visuals by Localizing Faces and Text Elements		
Direct all correspondence to : CORRESPONDENCE ADDRESS		
<input checked="" type="checkbox"/> Customer Number <input style="width: 150px;" type="text" value="000029690"/>	→	 <small>Patent Number Bar Code Label here</small>
OR		
ENCLOSED APPLICATION PARTS (check all that apply) <small>PATENT TRADEMARK OFFICE</small>		
<input checked="" type="checkbox"/> Specification	Number of Pages <u>18</u>	<input type="checkbox"/> CD(s), Number <input style="width: 50px;" type="text"/>
<input type="checkbox"/> Drawing(s)	Number of Sheets <input style="width: 50px;" type="text"/>	<input checked="" type="checkbox"/> Other (specify) <input style="width: 150px;" type="text" value="Postcard"/>
<input type="checkbox"/> Application Data Sheet, See 37 CFR 1.76		
METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT		
<input checked="" type="checkbox"/> Applicant claims small entity status. See 37 CFR 1.27.		FILING FEE AMOUNT (\$)
<input type="checkbox"/> A check or money order is enclosed to cover the filing fees.		<input style="width: 50px;" type="text" value="105"/>
<input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge filing Fees, any additional fees due or credit any overpayment to Deposit Account Number:	<input style="width: 100px;" type="text" value="50-1742"/>	
<input type="checkbox"/> Payment by credit card. Form PTO-2038 is attached.		
The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.		
<input checked="" type="checkbox"/> No		
<input type="checkbox"/> Yes, the name of the U.S. Government agency and the contract number are:		

Respectfully submitted,  Date

SIGNATURE _____

TYPED or PRINTED NAME: Frederic Farina

TELEPHONE: (626) 395-3058

REGISTRATION NO.
 (if appropriate)
 Docket Number:

Figure 144 - Provisional patent application

References

- Adolphs, R. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*(6507), 669-672.
- Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Sciences*, *3*(12), 469-479.
- Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, *12*(2), 169-177.
- Adolphs, R., Gosselin, F., Buchanan, T., Tranel, D., Schyns, P., & Damasio, A. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, *433*, 68-72.
- Adolphs, R., Spezio, M., Parlier, M., & Piven, J. (2008). Distinct Face-Processing Strategies in Parents of Autistic Children. *Current Biology*, *18*(14), 1090-1093.
- All, W. (1980). *Diagnostic and statistical manual of mental disorders*: American Psychiatric Association, Washington, DC.
- Allport, D. (1980). Attention and performance. *Cognitive psychology: New directions*, 112-153.
- Amaral, D. (2002). The primate amygdala and the neurobiology of social behavior: implications for understanding social anxiety. *Biological Psychiatry*, *51*(1), 11-17.
- Badaruddin, D., Andrews, G., Bölte, S., Schilmoeller, K., Schilmoeller, G., Paul, L., et al. (2007). Social and behavioral problems of children with agenesis of the corpus callosum. *Child Psychiatry and Human Development*, *38*(4), 287-302.

- Baird, G., Cass, H., & Slonims, V. (2003). Diagnosis of autism (Vol. 327, pp. 488-493): BMJ Publishing Group Ltd.
- Bar, M., Kassam, K., Ghuman, A., Boshyan, J., Schmid, A., Dale, A., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, *103*(2), 449.
- Barbeau, E., Taylor, M., Regis, J., Marquis, P., Chauvel, P., & Liegeois-Chauvel, C. (2008). Spatio temporal dynamics of face recognition. *Cerebral Cortex*, *18*(5), 997.
- Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences*, *6*(6), 248-254.
- Baron-Cohen, S. (2006). The hyper-systemizing, assortative mating theory of autism. *Progress in Neuropsychopharmacology & Biological Psychiatry*, *30*(5), 865-872.
- Baron-Cohen, S., Leslie, A., & Frith, U. (1985). Does the autistic child have a "theory of mind". *Cognition*, *21*(1), 13-125.
- Barton, J. (2003). Disorders of face perception and recognition. *Neurologic Clinics*, *21*(2), 521-548.
- Basso, M., & Wurtz, R. (1998). Modulation of neuronal activity in superior colliculus by changes in target probability. *Journal of Neuroscience*, *18*(18), 7519.
- Bauer, R. (1984). Autonomic recognition of names and faces in prosopagnosia: A neuropsychological application of the guilty knowledge test. *Neuropsychologia*, *22*(4), 457-469.
- Beaudet, A., & Zoghbi, H. (2006). A mixed epigenetic and genetic and mixed de novo and inherited model for autism. *Understanding Autism*, SO, Moldin, and JL, Rubenstein, eds. (Boca Raton: Taylor & Francis Group, LCC), 95-111.

- Beck, D., & Kastner, S. (2005). Stimulus context modulates competition in human extrastriate cortex. *Nature Neuroscience*, *8*, 1110-1116.
- Beck, D., & Kastner, S. (2007). Stimulus similarity modulates competitive interactions in human visual cortex. *Journal of Vision*, *7*(2), 19.
- Beck, D., & Kastner, S. (2008). Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Research*.
- Bedeschi, M., Bonaglia, M., Grasso, R., Pellegri, A., Garghentino, R., Battaglia, M., et al. (2006). Agenesis of the corpus callosum: clinical and genetic study in 63 young patients. *Pediatric Neurology*, *34*(3), 186-193.
- Ben Shalom, D. (2003). Memory in autism: review and synthesis. *Cortex*, *39*(4-5), 1129-1138.
- Benbadis, S., Gerson, W., Harvey, J., & Luders, H. (1996). Photosensitive temporal lobe epilepsy (Vol. 46, pp. 1540-1542): AAN Enterprises.
- Bennett, A., & Godlee, R. (1884). Excision of a tumour from the brain. *Lancet*, *2*, 1090-1091.
- Berger, H. (1929). On the electroencephalogram of man. *Journal fuer Psychiatric und Neurologic*, *28*, 37.
- Bettelheim, B. (1972). *Empty Fortress*: Free Press.
- Biederman, I. (1981). On the semantics of a glance at a scene. In M. K. J. R. Pomerantz (Ed.), *Perceptual Organization* (pp. 213-263): Lawrence Erlbaum, Hillsdale, New Jersey.
- Bindemann, M., Burton, A., Hooge, I., Jenkins, R., & de Haan, E. (2005). Faces retain attention. *Psychonomic Bulletin & Review*, *12*(6), 1048-1053.
- Bindemann, M., Burton, A., Langton, S., Schweinberger, S., & Doherty, M. (2007). The control of attention to faces. *Journal of Vision*, *7*(10), 15.

- Bisley, J., & Goldberg, M. (2003). Neuronal activity in the lateral intraparietal area and spatial attention. *Science*, *299*(5603), 81.
- Bliss, T., & Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of physiology*, *232*(2), 331-356.
- Bonnel, A., Stein, J., & Bertucci, P. (1992). Does attention modulate the perception of luminance changes? *The Quarterly Journal of Experimental Psychology. A - Human Experimental Psychology*, *44*(4), 601.
- Bonora, E., Lamb, J., Barnby, G., Bailey, A., & Monaco, A. (2006). Genetic Basis of Autism. *Understanding Autism, SO Moldin, JL, Rubenstein eds.,(Boca Raton, CRC Press, Taylor and Francis Group)*, 49-74.
- Bradski, G., Kaehler, A., & Pisarevsky, V. (2005). Learning-based computer vision with Intel's open source computer vision library. *Intel Technology Journal*, *9*(1).
- Brainard, D. H. (1996). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433-436.
- Breazeal, C., & Scassellati, B. (1999). A context-dependent attention system for a social robot. *International Joint Conference on Artificial Intelligence*, *16*, 1146-1153.
- Brefczynski, J., & DeYoe, E. (1999). A physiological correlate of the "spotlight" of visual attention. *Nature Neuroscience*, *2*, 370-374.
- Broadbent, D. (1958). *Perception and communication*: Pergamon Press.
- Brown, W., Jeeves, M., Dietrich, R., & Burnison, D. (1999). Bilateral field advantage and evoked potential interhemispheric transmission in commissurotomy and callosal agenesis. *Neuropsychologia*, *37*(10), 1165-1180.

- Bruce, V., & Langton, S. (1994). The use of pigmentation and shading information in recognising the sex and identities of faces. *Perception, 23*, 803-803.
- Bruce, V., & Young, A. (1986). Understanding face recognition. *Journal of Psychology, 77*(Pt 3), 305-327.
- Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews, 36*(2-3), 96-107.
- Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97*(4), 523.
- Busetini, C., Masson, G., & Miles, F. (1997). Radial optic flow induces vergence eye movements with ultra-short latencies. *Nature, 512-514*.
- Buswell, G. (1935). *How People Look at Pictures: A Study of the Psychology of Perception in Art*: The University of Chicago press.
- Calder, A., Young, A., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of experimental psychology: Human perception and performance, 26*(2), 527.
- Carmena, J., Lebedev, M., Crist, R., O'Doherty, J., Santucci, D., Dimitrov, D., et al. (2003). Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biology, 1*(2), e2.
- Carpenter, R., & Williams, M. (1995). Neural computation of log likelihood in control of saccadic eye movements. *nature, 377*(6544), 59-62.
- Cashon, C., & Cohen, L. (2003). The construction, deconstruction, and reconstruction of infant face perception. *The development of face processing in infancy and early childhood: Current perspectives, 55-68*.

- Cerf, M., Cleary, D., Peters, R., Einhäuser, W., & Koch, C. (2007). Observers are consistent when rating image conspicuity. *Vision Research*, *47*(24), 3052-3060.
- Cerf, M., Frady, E., & Koch, C. (2008). Using semantic content as cues for better scanpath prediction. *Proceedings of the 2008 symposium on Eye tracking research & applications*, 143-146.
- Cerf, M., Frady, E. P., & Koch, C. (2009). Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of Vision*, *9*(12), 1-15.
- Cerf, M., Harel, J., Einhäuser, W., & Koch, C. (2008). Predicting human gaze using low-level saliency combined with face detection. *Advances in Neural Information Processing Systems*, *20*.
- Cerf, M., Harel, J., Huth, A., W. E., & Koch, C. (2008). Decoding what people see from where they look: predicting visual stimuli from scanpaths. *International Workshop on Attention and Performance in Computational Vision*.
- Changizi, M., Zhang, Q., Ye, H., & Shimojo, S. (2006). The Structures of Letters and Symbols throughout Human History Are Selected to Match Those Found in Objects in Natural Scenes. *Am Nat*, *167*(5), E117-139.
- Chiarello, C. (1980). A house divided? Cognitive functioning with callosal agenesis. *Brain and language*, *11*(1), 128.
- Clow, K., & Baack, D. (2006). *Integrated advertising, promotion, and marketing communications*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- Cohen, L., Jobert, A., Le Bihan, D., & Dehaene, S. (2004). Distinct unimodal and multimodal regions for word processing in the left temporal cortex. *Neuroimage*, *23*(4), 1256-1270.

- Colby, C., Duhamel, J., & Goldberg, M. (1996). Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *Journal of Neurophysiology*, *76*(5), 2841-2852.
- Connor, C., Egeth, H., & Yantis, S. (2004). Visual attention: bottom-up versus top-down. *Current Biology*, *14*(19), 850-852.
- Cook Jr., E. (1998). Genetics of autism. *Mental Retardation and Developmental Disabilities Research Reviews*, *4*(2).
- Cornelissen, F., Peters, E., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with Matlab and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, *34*(4), 613-617.
- Dakin, S., & Frith, U. (2005). Vagaries of visual perception in autism. *Neuron*, *48*(3), 497-507.
- Darwin, C. (1872). The expression of the emotions in man and animals. *Murray, London*.
- Davenport, T., & Beck, J. (2001). *The attention economy: Understanding the new currency of business*: Harvard Business School Press.
- David, A., Wacharasindhu, A., & Lishman, W. (1993). Severe psychiatric disturbance and abnormalities of the corpus callosum: review and case series. *British Medical Journal*, *56*(1), 85-93.
- Deco, G., & Schurmann, B. (2000). A hierarchical neural system with attentional top-down enhancement of the spatial resolution for object recognition. *Vision Research*, *40*(20), 2845-2859.
- Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *353*(1373), 1245-1255.

- Desimone, R., Albright, T., Gross, C., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience*, *4*(8), 2051-2062.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, *18*(1), 193-222.
- Desimone, R., & Ungerleider, L. (1989). Neural mechanisms of visual processing in monkeys. *Handbook of neuropsychology*, *2*, 267-299.
- Deutsch, J., & Deutsch, D. (1963). Some theoretical considerations. *Psychological Review*, *70*, 80.
- Dickinson, S., Christensen, H., Tsotsos, J., & Olofsson, G. (1994). Active Object Recognition Integrating Attention and Viewpoint Control. *Computer Vision, ECCV'94: Proceedings of the Third European Conference on Computer Vision, Stockholm, Sweden, May 2-6, 1994*.
- Duchaine, B. (2000). Developmental prosopagnosia with normal configural processing. *Neuroreport*, *11*(1), 79.
- Duchaine, B., & Nakayama, K. (2005). Dissociations of face and object recognition in developmental prosopagnosia. *Journal of Cognitive Neuroscience*, *17*(2), 249-261.
- Duchaine, B., Parker, H., & Nakayama, K. (2003). Normal recognition of emotion in a prosopagnosic. *Perception*, *32*(7), 827-838.
- Duchowski, A. (2007). *Eye tracking methodology: Theory and practice*: Springer-Verlag Inc., New York.
- Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, *87*(3), 272-300.

- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of experimental psychology. General*, *113*(4), 501-517.
- Duncan, J. (1985). Visual search and visual attention. *Attention and performance*, *11*, 85-105.
- Duncan, J. (1993). Similarity between concurrent visual discriminations: Dimensions and objects. *Perception and Psychophysics*, *54*, 425.
- Duncan, J. (1996). Cooperating brain systems in selective perception and action. *Attention and performance*, *16*, 549-578.
- Egidi, G., Nusbaum, H., & Cacioppo, J. (2007). Neuroeconomics: Foundational issues and consumer relevance.
- Eichenbaum, H., Yonelinas, A., & Ranganath, C. (2007). The medial temporal lobe and recognition memory.
- Eifuku, S., De Souza, W., Tamura, R., Nishijo, H., & Ono, T. (2004). Neuronal correlates of face identification in the monkey anterior temporal cortical areas. *Journal of Neurophysiology*, *91*(1), 358.
- Einhäuser, W., & König, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *European Journal of Neuroscience*, *17*(5), 1089-1097.
- Einhäuser, W., Rutishauser, U., Frady, E., Nadler, S., König, P., & Koch, C. (2006). The relation of phase noise and luminance contrast to overt attention in complex visual stimuli. *Journal of Vision*, *6*(11), 1148-1158.
- Einhäuser, W. R., U., Frady, E.P., Nadler, S., König, P., Koch, C. (2006). The Relation of Phase Noise and Luminance Contrast to Overt Attention in Complex Visual Stimuli. *6*(11), 1148-1158.

- Ekman, P. (1973). Cross-cultural studies of facial expression. *Darwin and facial expression: A century of research in review*, 169-222.
- Ekstrom, A., Kahana, M., Caplan, J., Fields, T., Isham, E., Newman, E., et al. (2003). Cellular networks underlying human spatial navigation. *Nature*, *425*(6954), 184-188.
- Ellis, H., Whitley, J., & Luauté, J. (1994). Delusional misidentification: The three original papers on the Capgras, Fregoli and intermetamorphosis delusions. *History of Psychiatry*, *5*(17), 117.
- Engel, J. (1987). Surgical treatment of the epilepsies.
- Eriksen, C., & Hoffman, J. (1973). The extent of processing of noise elements during selective encoding from visual displays. *Perception & Psychophysics*, *14*(1), 155-160.
- Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology - Human Perception and Performance*, *31*(6), 1476-1492.
- Felleman, D., & Van Essen, D. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex*, *1*(1), 1-47.
- Fetz, E. (1969). Operant conditioning of cortical unit activity. *Science*, *163*(3870), 955.
- Fischer, M., Ryan, S., & Dobyns, W. (1992). Mechanisms of interhemispheric transfer and patterns of cognitive function in acallosal patients of normal intelligence. *Archives of Neurology*, *49*(3), 271-277.
- Fletcher-Watson, S., Findlay, J., Leekam, S., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception*, *37*, 571-583.

- Förstl, H., Almeida, O., Owen, A., Burns, A., & Howard, R. (1991). Psychiatric, neurological and medical aspects of misidentification syndromes: a review of 260 cases. *Psychological medicine(Print)*, *21*(4), 905-910.
- Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2), 6.
- Freedman, D., Riesenhuber, M., Poggio, T., & Miller, E. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *Journal of Neuroscience*, *23*(12), 5235.
- Freund, Y., Schapire, R., & Abe, N. (1999). A short introduction to boosting. *Japanese Society for Artificial Intelligence*, *14*, 771-780.
- Fried, I. (1993). Anatomic temporal lobe resections for temporal lobe epilepsy. *Neurosurg Clin N Am*, *4*(2), 233-242.
- Fried, I., MacDonald, K., & Wilson, C. (1997). Single Neuron Activity in Human Hippocampus and Amygdala during Recognition of Faces and Objects. *Neuron*, *18*, 753-765.
- Fujii, N., Mushiake, H., & Tanji, J. (1998). Intracortical microstimulation of bilateral frontal eye field. *Journal of Neurophysiology*, *79*(4), 2240.
- Galaburda, A., & Duchaine, B. (2003). Developmental disorders of vision. *Neurologic Clinics*, *21*(3), 687-707.
- Gauthier, I., Skudlarski, P., Gore, J., & Anderson, A. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*(2), 191-197.
- Gazzaniga, M. (1968). *The Split Brain in Man*: WH Freeman.

- Gazzaniga, M. (1970). The bisected brain.
- Golarai, G., Ghahremani, D., Whitfield-Gabrieli, S., Reiss, A., Eberhardt, J., Gabrieli, J., et al. (2007). Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nature Neuroscience*, *10*, 512-522.
- Goldstein, R., Woods, R., & Peli, E. (2007). Where people look when watching movies: Do all viewers look at the same place? *Computers in Biology and Medicine*, *37*(7), 957-964.
- Gross, C., & Schonen, S. (1992). Representation of Visual Stimuli in Inferior Temporal Cortex [and Discussion]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *335*(1273), 3-10.
- Gur, R., Sara, R., Hagendoorn, M., Marom, O., Huggett, P., Macy, L., et al. (2002). A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. *Journal of neuroscience methods*, *115*(2), 137-143.
- Hanes, D., & Schall, J. (1996). Neural control of voluntary movement initiation. *Science*, *274*(5286), 427.
- Happé, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*(1), 5-25.
- Harel, J., Koch, C., & Perona, P. (2007). Graph-Based Visual Saliency. *Advances in Neural Information Processing Systems*, *19*, 545.
- Harper, D. (2001). Online etymology dictionary. Retrieved Dec, 27, 2007.
- Hasselmo, M., Rolls, E., & Baylis, G. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioural Brain Research*, *32*(3), 203-218.

- Haxby, J., Hoffman, E., & Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6), 223-232.
- Henderson, J., Brockmole, J., Castelano, M., Mack, M., van Gompel, R., Fischer, M., et al. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. *Eye movements: A window on mind and brain*, 537-562.
- Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Research*, 45(13), 1707-1724.
- Hill, E. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences*, 8(1), 26-32.
- Hochberg, L., Serruya, M., Friehs, G., Mukand, J., Saleh, M., Caplan, A., et al. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442, 164-171.
- Honey, C., Kirchner, H., & VanRullen, R. (2008). Faces in the cloud: Fourier power spectrum biases ultrarapid face.
- Hopfinger, J., Buonocore, M., & Mangun, G. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience*, 3, 284-291.
- Hung, C., Kreiman, G., Poggio, T., & DiCarlo, J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, 310(5749), 863.
- Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience*, 7(12), 942-951.
- Imamura, T., Yamadori, A., Shiga, Y., Sahara, M., & Abiko, H. (1994). Is disturbed transfer of learning in callosal agenesis due to a disconnection syndrome? *Behavioural neurology*, 7(2), 43-48.

- Itti, L. (2004). Modeling primate visual attention. *Computational Neuroscience: A comprehensive approach*, 635-655.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12), 1489-1506.
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Review Neuroscience*, 2(3), 194-203.
- Itti, L., Koch, C., & Niebur, E. (1998). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1254-1259.
- James, W. (1890). *The Principles of Psychology*. New York: Henry Holt.
- Janiszewski, C. (1993). Preattentive mere exposure effects. *Journal of Consumer Research*, 376-392.
- Janiszewski, C. (1998). The influence of display characteristics on visual exploratory search behavior. *Journal of Consumer Research*, 25(3), 290-301.
- Johnson, M., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, 40(1-2), 1-19.
- Kamitani, Y., & Shimojo, S. (1999). Manifestation of scotomas created by transcranial magnetic stimulation of human visual cortex. *Nature Neuroscience*, 2, 767-771.
- Kandel, E., Schwartz, J., & Jessell, T. (2000). *Principles of neural science*: Appleton & Lange.
- Kant, I. (1882). *Critique of pure reason*: G. Bell.
- Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, 3, 759-763.

- Kanwisher, N., McDermott, J., & Chun, M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302-4311.
- Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science*, *282*(5386), 108.
- Kastner, S., De Weerd, P., Pinsk, M., Elizondo, M., Desimone, R., & Ungerleider, L. (2001). Modulation of sensory suppression: implications for receptive field sizes in the human visual cortex. *Journal of Neurophysiology*, *86*(3), 1398-1411.
- Kayser, C., Nielsen, K., & Logothetis, N. (2006). Fixations in natural scenes: Interaction of image structure and image content. *Vision Research*, *46*(16), 2535-2545.
- Kenworthy, L., Yerys, B., Anthony, L., & Wallace, G. (2008). Understanding Executive Control in Autism Spectrum Disorders in the Lab and in the Real World. *Neuropsychology Review*, *18*(4), 320-338.
- Keysers, C., Xiao, D., Földiák, P., & Perrett, D. (2001). The speed of sight. *Journal of cognitive neuroscience*, *13*(1), 90-101.
- Kim, S., Simeral, J., Hochberg, L., Donoghue, J., & Black, M. (2008). Neural control of computer cursor velocity by decoding motor cortical spiking activity in humans with tetraplegia. *Journal of Neural Engineering*, *5*(4), 455-476.
- Kirchner, H., Barbeau, E., Thorpe, S., Regis, J., & Liegeois-Chauvel, C. (2009). Ultra-rapid sensory responses in the human frontal eye field region. *Journal of Neuroscience*, *29*(23), 7599.

- Kirchner, H., & Thorpe, S. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, *46*(11), 1762-1776.
- Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual Fixation Patterns During Viewing of Naturalistic Social Situations as Predictors of Social Competence in Individuals With Autism (Vol. 59, pp. 809-816): Am Med Assoc.
- Koch, C. (2004). *The quest for consciousness: A neurobiological approach*. Roberts & Co.
- Koch, C., & Ullman, S. (1985). Shifts in Selective Visual-Attention - Towards the Underlying Neural Circuitry. *Human Neurobiology*, *4*(4), 219-227.
- Kraskov, A., Quiroga, R., Reddy, L., Fried, I., & Koch, C. (2007). Local field potentials and spikes in the human medial temporal lobe are selective to image category. *Journal of Cognitive Neuroscience*, *19*(3), 479-492.
- Kreiman, G. (2001). *On the neuronal activity in the human brain during visual recognition, imagery and binocular rivalry*. California Institute of Technology.
- Kreiman, G., Hung, C., Kraskov, A., Quiroga, R., Poggio, T., & DiCarlo, J. (2006). Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex. *Neuron*, *49*(3), 433-445.
- Kreiman, G., Koch, C., & Fried, I. (2000a). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature Neuroscience*, *3*, 946-953.
- Kreiman, G., Koch, C., & Fried, I. (2000b). Imagery neurons in the human brain. *Nature*, *408*(6810), 357-361.
- Krolak-Salmon, P., Henaff, M., Tallon-Baudry, C., Yvert, B., Guénot, M., Vighetto, A., et al. (2003). Human lateral geniculate nucleus and visual cortex respond to screen flicker. *Annals of neurology*, *53*(1), 73-80.

- Lamb, J., Moore, J., Bailey, A., & Monaco, A. (2000). Autism: recent molecular genetic advances (Vol. 9, pp. 861-868): Oxford Univ Press.
- Lang, P., Bradley, M., & Cuthbert, B. (1995). International affective picture system (IAPS): Technical manual and affective ratings. *The Center for Research in Psychophysiology, University of Florida*.
- Lassonde, M., & Jeeves, M. (1994). *Callosal agenesis: a natural split brain?*: Plenum Press.
- Ledberg, A., Bressler, S., Ding, M., Coppola, R., & Nakamura, R. (2007). Large-scale visuomotor integration in the cerebral cortex. *Cerebral Cortex*, *17*(1), 44.
- Leopold, D., O'Toole, A., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, *4*, 89-94.
- Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center-periphery organization of human object areas. *Nature Neuroscience*, *4*(5), 533-539.
- Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14), 9596-9601.
- Liu, H., Agam, Y., Madsen, J., & Kreiman, G. (2009). Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron*, *62*(2), 281-290.
- Logothetis, N., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *CURRENT BIOLOGY*, *5*, 552-552.
- Lord, C., Rutter, M., & Couteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive

- developmental disorders. *Journal of Autism and Developmental Disorders*, 24(5), 659-685.
- Luck, S., Chelazzi, L., Hillyard, S., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology*, 77(1), 24-42.
- Mack, A., Pappas, Z., Silverman, M., & Gay, R. (2002). What we see: Inattention and the capture of attention by meaning. *Consciousness and Cognition*, 11(4), 488-506.
- McKone, E., Kanwisher, N., & Duchaine, B. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences*, 11(1), 8-15.
- Mishkin, M., Ungerleider, L., & Macko, K. (2001). Object vision and spatial vision: Two cortical pathways. *Philosophy and the Neurosciences: A Reader*, 199.
- Mormann, F., Kornblith, S., Quiroga, R., Kraskov, A., Cerf, M., Fried, I., et al. (2008). Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. *Journal of Neuroscience*, 28(36), 8865.
- Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*, 36(1), 27-43.
- Moutard, M., Kieffer, V., Feingold, J., Kieffer, F., Lewin, F., Adamsbaum, C., et al. (2003). Agenesis of corpus callosum: prenatal diagnosis and prognosis. *Child's Nervous System*, 19(7), 471-476.
- Munk, M., Nowak, L., Girard, P., Chounlamountri, N., & Bullier, J. (1995). Visual latencies in cytochrome oxidase bands of macaque area V2. *Proceedings of the National Academy of Sciences*, 92(4), 988.

- Munoz, D., & Wurtz, R. (1993). Fixation cells in monkey superior colliculus. I. Characteristics of cell discharge. *Journal of Neurophysiology*, *70*(2), 559-575.
- Musallam, S., Andersen, R., Corneil, B., Greger, B., & Scherberger, H. (2005). Cognitive control signals for neural prosthetics: Google Patents.
- Musallam, S., Corneil, B., Greger, B., Scherberger, H., & Andersen, R. (2004). Cognitive control signals for neural prosthetics. *Science*, *305*(5681), 258-262.
- Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, *45*(2), 205-231.
- Navalpakkam, V., & Itti, L. (2007). Search Goal Tunes Visual Features Optimally. *Neuron*, *53*(4), 605-617.
- Naya, Y., Yoshida, M., & Miyashita, Y. (2001). Backward spreading of memory-retrieval signal in the primate temporal cortex. *Science*, *291*(5504), 661.
- Naya, Y., Yoshida, M., & Miyashita, Y. (2003). Forward processing of long-term associative memory in monkey inferotemporal cortex. *Journal of Neuroscience*, *23*(7), 2861.
- Neisser, U. (1967). *Cognitive psychology*: Appleton-Century-Crofts.
- O'Craven, K., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*(6), 1013-1023.
- O'Hearn, K., Asato, M., Ordaz, S., & Luna, B. (2008). Neurodevelopment and executive function in autism. *Development and Psychopathology*, *20*(04), 1103-1132.
- Ogawa, T., & Komatsu, H. (2004). Target selection in area V4 during a multidimensional visual search task. *Journal of Neuroscience*, *24*(28), 6371.

- Ojemann, G., & Silbergeld, D. (1993). Approaches to epilepsy surgery. *Neurosurg Clin N Am*, 4(2), 183-191.
- Oliva, A., & Torralba, A. (2006). Building the Gist of a Scene: The Role of Global Image Features in Recognition. *Progress in Brain Research*.
- Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, 11(12), 520-527.
- Oliva, A., Torralba, A., Castelano, M., & Henderson, J. (2003). Top-down control of visual attention in object detection. *Proceedings of the International Conference on Image Processing*, 1.
- Oram, M., & Perrett, D. (1992). Time course of neural responses discriminating different views of the face and head. *Journal of Neurophysiology*, 68(1), 70.
- Pack, C., & Born, R. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *nature*, 409(6823), 1040-1042.
- Palermo, R., & Rhodes, G. (2007). Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia*, 45(1), 75-92.
- Paré, M., & Munoz, D. (1996). Saccadic reaction time in the monkey: advanced preparation of oculomotor programs is primarily responsible for express saccade occurrence. *Journal of Neurophysiology*, 76(6), 3666.
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42(1), 107-123.
- Pascual-Leone, A., Tormos, J., Keenan, J., Tarazona, F., Cañete, C., & Catalá, M. (1998). Study and modulation of human cortical excitability with transcranial magnetic stimulation. *Journal of Clinical Neurophysiology*, 15(4), 333.

- Paul, L., Brown, W., Adolphs, R., Tyszka, J., Richards, L., Mukherjee, P., et al. (2007). Agenesis of the corpus callosum: genetic, developmental and functional aspects of connectivity. *Nature Reviews Neuroscience*, *8*(4), 287-299.
- Paul, L., Schieffer, B., & Brown, W. (2004). Social processing deficits in agenesis of the corpus callosum: narratives from the Thematic Apperception Test. *Archives of Clinical Neuropsychology*, *19*(2), 215-225.
- Pearson, D., Hanna, E., & Martinez, K. (1990). Computer-generated cartoons. *Images and understanding*, 46-60.
- Perrett, D., Rolls, E., & Caan, W. (1982a). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, *47*(3), 329-342.
- Persico, A., & Bourgeron, T. (2006). Searching for ways out of the autism maze: genetic, epigenetic and environmental clues. *Trends in Neurosciences*, *29*(7), 349-358.
- Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences*, *99*(17), 11458-11463.
- Peters, R., & Itti, L. (2007). Congruence between model and human attention reveals unique signatures of critical visual events. *Advances in Neural Information Processing Systems*, *20*, 1145-1152.
- Peters, R., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, *45*(18), 2397-2416.
- Pieters, R., & Wedel, M. (2004). Attention capture and transfer in advertising: brand, pictorial, and text-size effects. *Journal of Marketing*, *68*(2), 36-50.

- Pieters, R., & Wedel, M. (2007). Goal control of attention to advertising: The Yarbus implication. *Journal of Consumer Research*, *34*(2), 224-233.
- Plassmann, H., Ambler, T., Braeutigam, S., & Kenning, P. (2007). What can advertisers learn from neuroscience? *International Journal of Advertising*, *26*(2), 151.
- Posner, M., & Petersen, S. (1990). The attention system of the human brain. *Annual review of neuroscience*, *13*(1), 25-42.
- Posner, M., Snyder, C., & Davidson, B. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160-174.
- Potter, M. C. (1976). Short-Term Conceptual Memory for Pictures. *Journal of Experimental Psychology-Human Learning and Memory*, *2*(5), 509-522.
- Potter, M. C., & Levy, E. I. (1969). Recognition Memory for a Rapid Sequence of Pictures. *Journal of Experimental Psychology*, *81*(1), 10.
- Potter, M. C., Staub, A., Rado, J., & O'Connor, D. H. (2002). Recognition memory for briefly presented pictures: The time course of rapid forgetting. *Journal of Experimental Psychology-Human Perception and Performance*, *28*(5), 1163-1175.
- Pring, L. (2005). Savant talent. *Developmental Medicine and Child Neurology*, *47*(07), 500-503.
- Quiroga, R., Mukamel, R., Isham, E., Malach, R., & Fried, I. (2008). Human single-neuron responses at the threshold of conscious recognition. *Proceedings of the National Academy of Sciences*, *105*(9), 3599.
- Quiroga, R., Nadasdy, Z., & Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural computation*, *16*(8), 1661-1687.

- Quiroga, R., Reddy, L., Koch, C., & Fried, I. (2007). Decoding Visual Inputs from Multiple Neurons in the Human Temporal Lobe. *Journal of Neurophysiology*, *98*(4), 1997.
- Quiroga, R., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*(7045), 1102-1107.
- Raiguel, S., Lagae, L., Guly s, B., & Orban, G. (1989). Response latencies of visual cells in macaque areas V1, V2 and V5. *Brain research*, *493*(1), 155-159.
- Raiguel, S., Xiao, D., Marcar, V., & Orban, G. (1999). Response latency of macaque area MT/V5 neurons and its relationship to stimulus parameters. *Journal of Neurophysiology*, *82*(4), 1944.
- Raybourn, M., & Keller, E. (1977). Colliculoreticular organization in primate oculomotor system. *J Neurophysiol*, *40*(4), 861-878.
- Rayner, K., Rotello, C., Stewart, A., Keir, J., & Duffy, S. (2001). Integrating text and pictorial information: eye movements when looking at print advertisements. *Journal of experimental psychology. Applied*, *7*(3), 219.
- Reddi, B. (2001). Decision making: The two stages of neuronal judgement. *Current Biology*, *11*(15), 603-606.
- Reddi, B., Asress, K., & Carpenter, R. (2003). Accuracy, information, and response time in a saccadic decision task. *Journal of Neurophysiology*, *90*(5), 3538.
- Reddi, B., & Carpenter, R. (2000). The influence of urgency on decision time. *nature neuroscience*, *3*, 827-830.
- Reddy, L., Wilken, P., & Koch, C. (2004). Face-gender discrimination is possible in the near-absence of attention. *Journal of Vision*, *4*(2), 106-117.

- Renninger, L. W., & Malik, J. (2004). When is scene identification just texture recognition? *Vision Research*, *44*(19), 2301-2311.
- Reynolds, J., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience*, *19*(5), 1736-1753.
- Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nat Neurosci*, *3 Suppl*, 1199-1204.
- Rimland, B. (1964). *Infantile autism: The syndrome and its implications for a neural theory of behavior*. Appleton-Century-Crofts.
- Rizzolatti, G., Riggio, L., Dascola, I., & Umiltà, C. (1987). Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. *Neuropsychologia*, *25*(1A), 31-40.
- Ro, T., Russell, C., & Lavie, N. (2001). Changing Faces: A Detection Advantage in the Flicker Paradigm. *Psychological Science*, *12*(1), 94-99.
- Rosbergen, E., Pieters, R., & Wedel, M. (1997). Visual attention to advertising: A segment-level analysis. *Journal of Consumer Research*, *24*(3), 305-314.
- Rosenow, F., & Luders, H. (2001). Presurgical evaluation of epilepsy. *Brain*, *124*(9), 1683.
- Roth, D., Yang, M., & Ahuja, N. (2000). A snowbased face detector. *Neural Information Processing*, *12*.
- Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience*, *5*(7), 629-630.
- Rubin, E. (1958). Figure and ground. *In Readings in Perception: Selected and Edited by David C. Beardslee and Michael Wertheimer*, 194.

- Ruohonen, J. (1998). *Transcranial magnetic stimulation: Modelling and new techniques*. J. Ruohonen.
- Rutter, M. (1968). Concepts of Autism: A review of research. *Journal of Child Psychology and Psychiatry*, 9(1), 1-25.
- Sacks, O. (1995). The man who mistook his wife for a hat. *The British Journal of Psychiatry*, 166(1), 130.
- Sagi, D., & Julesz, B. (1985). Fast noninertial shifts of attention. *Spatial Vision*, 1(2), 141-149.
- Sanders, J., Johnson, K., Garavan, H., Gill, M., & Gallagher, L. (2008). A review of neuropsychological and neuroimaging research in autistic spectrum disorders: Attention, inhibition and cognitive flexibility. *Research in Autism Spectrum Disorders*, 2(1), 1-16.
- Sasson, N., Tsuchiya, N., Hurley, R., Couture, S., Penn, D., Adolphs, R., et al. (2007). Orienting to social stimuli differentiates social cognitive impairment in autism and schizophrenia. *Neuropsychologia*, 45(11), 2580-2588.
- Sato, T., Kawamura, T., & Iwai, E. (1980). Responsiveness of inferotemporal single units to visual pattern stimuli in monkeys performing discrimination. *Experimental Brain Research*, 38(3), 313-319.
- Schall, J., & Hanes, D. (1993). Neural basis of saccade target selection in frontal eye field during visual search. *nature*, 366(6454), 467-469.
- Schall, J., & Thompson, K. (1999). Neural selection and control of visually guided eye movements. *Annual review of neuroscience*, 22(1), 241-259.
- Schmolesky, M., Wang, Y., Hanes, D., Thompson, K., Leutgeb, S., Schall, J., et al. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology*, 79(6), 3272.

- Schwartz, A., Cui, X., Weber, D., & Moran, D. (2006). Brain-controlled interfaces: movement restoration with neural prosthetics. *Neuron*, *52*(1), 205-220.
- Schwartz, E., Desimone, R., Albright, T., & Gross, C. (1983). Shape Recognition and Inferior Temporal Neurons. *Proceedings of the National Academy of Sciences*, *80*(18), 5776-5778.
- Schyns, P. G., & Oliva, A. (1997). Flexible, diagnosticity-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. *Perception*, *26*(8), 1027-1038.
- Sejnowski, T., Koch, C., & Churchland, P. (1988). Computational neuroscience. *Science*, *241*(4871), 1299-1306.
- Semah, F. (1998). Is the underlying cause of epilepsy a major prognostic factor for recurrence? *Neurology*, *51*(5), 1256-1262.
- Shapiro, S., MacInnis, D., & Heckler, S. (1997). The effects of incidental ad exposure on the formation of consideration sets. *Journal of Consumer Research*, *24*(1), 94-104.
- Shapiro, S., MacInnis, D., Heckler, S., & Perez, A. (1999). An experimental method for studying unconscious perception in a marketing context. *Psychology and Marketing*, *16*(6), 459-477.
- Shevell, M. (2002). Clinical and diagnostic profile of agenesis of the corpus callosum. *Journal of child neurology*, *17*(12), 895.
- Shipp, S. (2004). The brain circuitry of attention. *Trends in Cognitive Sciences*, *8*(5), 223-230.
- Sigman, M., Spence, S., & Wang, A. (2006). Autism from developmental and neuropsychological perspectives.
- Simion, C., & Shimojo, S. (2006). Early interactions between orienting, visual sampling and decision making in facial preference. *Vision Research*, *46*(20), 3331-3335.

- Sinha, P., Balas, B., Ostrovsky, Y., & Russell, R. (2006). Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, *94*(11), 1948.
- Sinha, P., & Poggio, T. (1996). I think I know that face. *Nature*, *384*(6608), 404.
- Sirovich, L., & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, *4*(3), 519-524.
- Smith, E., & Kosslyn, S. (2007). *Cognitive psychology: mind and brain*. Prentice Hall.
- Smith, J., VanderGriff, A., & Fountas, K. (2004). Temporal lobectomy in the surgical management of epilepsy: technical report. *Neurosurgery*, *54*(6), 1531.
- Solursh, L., Psych, D., Margulies, A., Ashem, B., & Stasiak, E. (1965). The relationships of agenesis of the corpus callosum to perception and learning. *The Journal of Nervous and Mental Disease*, *141*(2), 180.
- Sommer, M. (1997). The spatial relationship between scanning saccades and express saccades. *Vision Research*, *37*(19), 2745-2756.
- Sparks, D. (1978). Functional properties of neurons in the monkey superior colliculus: coupling of neuronal activity and saccade onset. *Brain Res*, *156*(1), 1-16.
- Sperry, R., Schmitt, F., & Worden, F. (1974). *The neurosciences: Third study program*. MIT Press Cambridge.
- Standing, L. (1973). Learning 10,000 Pictures. *Quarterly Journal of Experimental Psychology*, *25*, 207-222.
- Stickles, J., Schilmoeller, G., & Schilmoeller, K. (2002). A 23-year review of communication development in an individual with agenesis of the corpus callosum. *International Journal of Disability, Development and Education*, *49*(4), 367-383.

- Sun, Y. R., & Fisher, R. (2003). Object-based visual attention for computer vision. *Artificial Intelligence, 146*(1), 77-123.
- Sung, K., & Poggio, T. (1998). Example-based learning for view-based human face detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 20*(1), 39-51.
- Tamura, H., & Tanaka, K. (2001). Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cerebral Cortex, 11*(5), 384.
- Tatler, B., Baddeley, R., & Gilchrist, I. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research, 45*(5), 643-659.
- Theeuwes, J., & Van der Stigchel, S. (2006). Faces capture attention: Evidence from inhibition of return. *Visual Cognition, 13*(6), 657-665.
- Thomas, N., & Pare, M. (2007). Temporal processing of saccade targets in parietal cortex area LIP during visual search. *Journal of Neurophysiology, 97*(1), 942.
- Thompson, K., Hanes, D., Bichot, N., & Schall, J. (1996). Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. *Journal of Neurophysiology, 76*(6), 4040.
- Thompson, P., Facial, E., Famous, P., & Optical, I. (1980). Margaret Thatcher: a new illusion. *Perception, 9*(4), 483-484.
- Thorpe, S., Delorme, A., & Van Rullen, R. (2001). Spike-based strategies for rapid processing. *Neural Networks, 14*(6-7), 715-725.
- Thorpe, S., & Fabre-Thorpe, M. (2001). Seeking categories in the brain. *Science, 260*-262.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *nature, 381*(6582), 520-522.

- Tong, F. (2000). Response properties of the human fusiform face area. *The Cognitive Neuroscience of Face Processing*, 2000(17), 257-279.
- Torralba, A., Oliva, A., Castelano, M., & Henderson, J. (2006). Contextual Guidance of Eye Movements and Attention in Real-World Scenes: The Role of Global Features in Object Search. *Psychological Review*, 113(4), 766.
- Tovee, M., Rolls, E., Treves, A., & Bellis, R. (1993). Information encoding and the responses of single neurons in the primate temporal visual cortex. *Journal of Neurophysiology*, 70(2), 640.
- Tranel, D., Damasio, H., & Damasio, A. (1995). Double dissociation between overt and covert face recognition. *Journal of Cognitive Neuroscience*, 7(4), 425-432.
- Treisman, A. (1969). Strategies and models of selective attention. *Psychological Review*, 76(3), 282.
- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, 12(1), 97-136.
- Treue, S., & Martinez-Trujillo, J. (2003). Cognitive Physiology: Moving the Mind's Eye before the Head's Eye. *Current Biology*, 13(11), 442-444.
- Tsao, D., Freiwald, W., Tootell, R., & Livingstone, M. (2006). A cortical region consisting entirely of face-selective cells. *Science*, 311(5761), 670-674.
- Tsao, D., Moeller, S., & Freiwald, W. (2008). Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences*, 105(49), 19514.
- Tsotsos, J. (1990). Analyzing vision at the complexity level. *Behavioral and brain sciences*, 13(3), 423-469.

- Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y. H., Davis, N., & Nuflo, F. (1995). Modeling Visual-Attention Via Selective Tuning. *Artificial Intelligence*, 78(1-2), 507-545.
- Turk, M., & Pentland, A. (1991). *Face recognition using eigenfaces*.
- Underwood, G., Foulsham, T., van Loon, E., & Underwood, J. (2005). Visual attention, visual saliency, and eye movements during the inspection of natural scenes. In *Artificial Intelligence and Knowledge Engineering Applications: a Bioinspired Approach, Pt 2, Proceedings* (Vol. 3562, pp. 459-468).
- Ungerleider, L., & Haxby, J. (1994). "What" and "where" in the human brain. *Current Opinion in Neurobiology*, 4(2), 157-165.
- Vakratsas, D., & Ambler, T. (1999). How advertising works: what do we really know? *The Journal of Marketing*, 26-43.
- Valentine, T., & Cross, N. (2001). Face-space models of face recognition. *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges*, 83-113.
- Valenzano, D., Mennucci, A., Tartarelli, G., & Cellerino, A. (2006). Shape analysis of female facial attractiveness. *Vision Research*, 46(8-9), 1282-1291.
- Van Essen, D., Anderson, C., & Felleman, D. (1992). Information processing in the primate visual system: an integrated systems perspective. *Science*, 255(5043), 419-423.
- Van Essen, D., & Maunsell, J. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends Neurosci*, 6(9), 370-375.
- van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 265(1394), 359-366.

- VanRullen, R. (2006). On second glance: Still no high-level pop-out effect for faces. *Vision Research*, *46*(18), 3017-3027.
- Vecera, S., & Farah, M. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology - General*, *123*(2), 146-160.
- Velliste, M., Perel, S., Spalding, M., Whitford, A., & Schwartz, A. (2008). Cortical control of a prosthetic arm for self-feeding. *Nature*, *453*(7198), 1098-1101.
- Viola, P., & Jones, M. (2001). Rapid Object Detection Using a Boosted Cascade of Simple Features. *IEEE Conference on Computer Vision and Pattern Recognition*, *1*.
- Volkmar, F., Chawarska, K., & Klin, A. (2004). Autism in infancy and early childhood.
- Vuilleumier, P. (2000). Faces call for attention: evidence from patients with visual extinction. *Neuropsychologia*, *38*(5), 693-700.
- Walther, D. (2006). *Interactions of visual attention and object recognition: computational modeling, algorithms, and psychophysics*.
- Walther, D., Koch, C. (2006). Modeling Attention to Salient Proto-Objects. *19*, 1395-1407.
- Watanabe, K., Lauwereyns, J., & Hikosaka, O. (2003). Neural correlates of rewarded and unrewarded eye movements in the primate caudate nucleus. *Journal of Neuroscience*, *23*(31), 10052.
- Waydo, S., Kraskov, A., Quiñ Quiroga, R., Fried, I., & Koch, C. (2006). Sparse representation in the human medial temporal lobe. *Journal of Neuroscience*, *26*(40), 10232.
- Webster, M., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, *428*, 557-561.
- Wechsler, D. (2007). *The measurement of adult intelligence*: Cooper Press.

- Wegner, D. (1994). *White bears and other unwanted thoughts*: Guilford Pr.
- Wessberg, J., Stambaugh, C., Kralik, J., Beck, P., Laubach, M., Chapin, J., et al. (2000). Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature*, *408*(6810), 361-365.
- Williams, D., Goldstein, G., & Minshew, N. (2005). Impaired memory for faces and social scenes in autism: clinical implications of memory dysfunction. *Archives of Clinical Neuropsychology*, *20*(1), 1-15.
- Wilson, F., Scaldie, S., & Goldman-Rakic, P. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science*, *260*(5116), 1955.
- Wolfe, J., & Horowitz, T. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Review Neuroscience*, *5*(6), 495-501.
- Wright, M. J. (2005). Saliency predicts change detection in pictures of natural scenes. *Spatial Vision*, *18*(4), 413-430.
- Yang, M., Kriegman, D., & Ahuja, N. (2002). Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(1), 34-58.
- Yarbus, A. (1967). *Eye Movements and Vision*: Plenum Press.
- Yin, R. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141-145.
- Yip, A., & Sinha, P. (2002). Role of color in face recognition. *Perception*, *31*, 995-1003.
- Yoo, C. (2005). *Preattentive processing of web advertising*: University of Texas at Austin Austin, TX, USA.
- Young, M., & Yamane, S. (1992). Sparse population coding of faces in the inferotemporal cortex. *Science*, *256*(5061), 1327-1331.

Zaltman, G. (2003). *How customers think: Essential insights into the mind of the market*.
Harvard Business School Press.

For 14 years, since I got a poster with that phrase on it the age of 17 in a short visit to Paris,
that poster hung on the wall, wherever I lived.

The poster had Einstein's quote that I reflected on whenever I accomplished a task that had to
do with increasing the amount of knowledge within myself, or in the world — as I hope I did in
this thesis.

This quote was never more important to me as it is when I complete this work.

Imagination is more important than knowledge.