RESEARCH STRATEGY

Significance

In this proposal, we focus on electronic (both retinal and cortical) and optogenetic sight recovery technologies. For all these technologies, interactions between implant technology and the underlying neurophysiology of the retina and cortex result in vision that differs substantially from normal sight[25]. Currently, device developers tend to optimistically pronounce that *'patients will learn to make use of their prosthetic vision'*. This claim is primarily motivated by the success of cochlear implants. Although the auditory signal from the earliest cochlear implants was too diminished to allow for speech perception[37], current cochlear implants provide most adult-deafened patients with surprisingly good speech recognition; there are rapid improvements within hours, followed by more gradual improvement over approximately a year[38].

However, plasticity may be very different for visual than for cochlear implants. Early stages of the visual hierarchy (V1) show far less plasticity than A1 or S1[39,40]. The reason for this is somewhat mysterious, but one possible explanation is that there is significantly more subcortical processing within somatosensory and auditory than within visual pathways. Thus, A1 and S1 can be considered as 'higher' in their respective processing pathways than V1 is within the visual processing hierarchy. Regardless of the cause, the restricted plasticity that is observed in V1 suggests that we should be cautious in our expectations when hoping that patients can learn to interpret the input from visual prosthetics.

Previous research examining the ability of individuals to make use of prosthetic vision have focused on the loss of resolution due to a limited number of electrodes[41-47], and have assumed that each electrode produces a well-behaved 'spot' of light. We call this assumption the 'pixelated scoreboard' model, since it resembles how images are generated using a matrix of lights on a stadium scoreboard. A prediction of the scoreboard model is that perfect (grayscale) vision should be achieved in the limit as electrodes become smaller and the field of view increases. Thus, under the scoreboard model, engineering is the limiting factor for sight restoration. The work proposed here is different because it focuses on more realistic spatiotemporal distortions caused by interactions between the technology and the underlying neurophysiology: *distortions that cannot be corrected simply by increasing the number of electrodes.*

Previous work has also demonstrated that individuals rapidly improve in their ability to compensate for low resolution pixelated vision[43], low pass filtered images[48], or external noise[49], and the mechanisms underlying these types of perceptual learning are fairly well understood. As described below, adaptation to the distortions simulated here are likely to rely on very different mechanisms, which have barely been explored.

We focus here on generating virtual patients for 'idealized' implants of extremely high resolution – a best-case scenario for future technology. Once we have characterized the limits of plasticity using these idealized devices, a goal for future funding periods will be to examine perceptual learning for neurophysiological distortions that are matched to devices implanted in actual patients.

Our goal is to train visually normal individuals using distorted input that simulates four of the critical distortions that will be inherent in sight restoration technologies within the foreseeable future. For each distortion, we will examine the extent to which our 'virtual patients' can adapt as a result of training. Training will be done in the context of bilateral visual deprivation and/or video game playing, since both methods have been shown to be surprisingly effective at eliciting plasticity, see *General Methods*.

There are, of course, major differences between any protocol that can be carried out in neurotypical sighted participants and the experience of sight recovery patients. One major difference is that sight recovery patients have access to distorted information for much more than 1 hour/day. However, it is worth noting that current Argus II retinal implant patients, by choice, tend to use their implant for only a couple of hours a day. The reason for this is unclear, but it seems plausible that the cognitive effort of decoding distorted input is a factor.

A second difference is that for real patients the alternative to distorted input is no input, whereas our virtual patients will spend most of their day with normal vision. This is likely to impact plasticity in two ways. (1) Deprivation has been shown to have dramatic effects on neurotransmitters associated with responsiveness and plasticity. Congenital blindness has been associated with a ~50 % reduction in GABA[50] (though see[51]). However, these effects of deprivation on neurotransmitter levels may be primarily mediated on a relatively short timescale[52-55]. Thus, it seems plausible that short term binocular deprivation can at least provide a weaker proxy for the effects of adult long-term deprivation. (2) Daily alternation with normal visual input may impair adaptation to distorted input. In macaques, training to detect electrical stimulation of the cortex causes a large,

reversible, retinotopically localized impairment of thresholds for detecting visual stimuli. Retraining on visual detection restores normal light thresholds, but at the cost of increased thresholds for detecting micro stimulation. These results naturally raise the concern that optimized decoding for electrical and light stimulation cannot simultaneously co-exist within a local cortical region[56]. However, the macaques were not trained under conditions that would be designed to promote generalization across the two types of input, and macaques are frustratingly notorious for failing to show generalization of learning under conditions where humans generalize effortlessly. Work done with prisms[57], colored lenses[58], and selective attenuation of certain orientations[59,60] suggests that humans are very capable of 'switching between perceptual modes', and of course any wearer of corrective lenses is similarly used to rapidly switching between modes of perceptual distortion.

Innovation

- Measurements of perceptual learning in the context of prosthetic simulations that include the effects of neurophysiological spatiotemporal distortions. The Argus I (Second Sight), the first retinal prosthesis to reach clinical trials (in 2003), was heavily based on the technology of a 16-electrode cochlear implant (Advanced Bionics). Clinical trials with this initial prototype immediately revealed significant perceptual spatiotemporal distortions, as extensively characterized by co-PI Fine[35,61-66]. Stimulating a single electrode rarely produced the experience of a 'dot' of light, instead leading to percepts that drastically varied in shape (described as 'blobs', 'streaks', or 'half-moons'). As a result, patients did not report seeing an interpretable world. The second-generation prototype, the Argus II, had 60 electrodes, and had a development cost estimated at ~200 million dollars, of which at least 10% was funded by the government. These spatiotemporal distortions (along with the difficulty in obtaining consistently optimal surgical placement) resulted in performance enhancements for patients being considerably less than might be expected from a scoreboard model, given the increase in the number of electrodes from 16 to 60. Indeed, the recognition that increasing the number of electrodes was likely to result in only marginal visual benefits played a significant role in Second Sight Medical Products' recent decision to halt production of the Argus II, curtail development of higher resolution devices, and focus on development of the Orion cortical implant.

- Currently, worldwide, there are over 42 companies/university groups that are actively developing retinal and cortical implants. The cost of all current sight restoration technologies is likely to surpass a billion dollars in the next decade, with a significant proportion funded by taxpayers. This will be the first work to systematically examine plasticity in response to the abnormal input resulting from prosthetic vision. Engineering decisions are still made by 'best-guessing' the complex interactions between electronics and neurophysiology. We will be the first to characterize these distortions, and determine which spatiotemporal distortions can be compensated for by plasticity, and which must be compensated for in device design.

- A new framework for conceptualizing plasticity. This proposal is equally innovative from a basic science perspective. The previous literature on perceptual learning and plasticity has mainly focused on two frameworks. The first examines how individuals learn to perceptually identify or discriminate a particular set of stimuli or tasks (e.g. the direction of a field of moving dots, identifying an object in noise[67] – the development of perceptual 'templates'. The second examines experiential (i.e. naturalistic viewing conditions) adaptation to sensory loss, for example, within a region of the visual field[68-73], within one eye[55,74,75], or by removing orientation[59,76], or spatial frequency[77,78] information. Our proposal frames the role of plasticity in a novel way: Is it possible to reconfigure the fundamental building blocks of visual perception in adults? This is a central question: both because of its translational importance, and because we will be the first to examine the adult-analogue of processes that are fundamental to early visual development.

- Realistic simulations of prosthetic performance. The prevailing 'pixelated' scoreboard' approach to simulating prosthetic vision is to assume that each electrode produces a 'dot of light' in the visual field location corresponding to the position of the electrode on the retina, whose brightness scales linearly with current amplitude[41,43]. Performance predictions resulting from such studies and simulations are misleading because they do not take account of the substantial distortions in space and time observed by actual patients. This lack of realism is a serious issue: it misleads the press, surgeons and patients. While our perceptual learning studies will focus on idealized devices, our simulations will also allow us to generate reasonable prosthetic simulations for a wide range of current devices. These simulations will (1) help manufacturers optimize physical device design and stimulation protocols, (2) help the FDA design tests that are suitable for assessing prosthetic performance, (3) guide rehabilitation specialists about how best to train patients to make use of devices, and (4)

guide the public, patient families and doctors towards realistic expectations about when an implant is likely to be helpful. An R00 to Michael Beyeler supports this optimization/simulation work for retinal implants. Cortical/optogenetic implant optimization simulations will be a likely additional byproduct of the work described in this proposal. Because simulations of specific devices are heavily contingent on current collaborations with device manufacturers, we have not included this work as a Specific Aim, thus our ability to complete our Aims is not dependent on collaborations with device manufacturers.

- Open science. The field of sight restoration is dominated by device manufacturers that own most of the intellectual property and control access to patients. While we have been very fortunate to be able to collaborate with Second Sight, Pixium Vision, CORTVis, and Vedere, they are commercial competitors. To avoid any conflict of interest and to accelerate scientific insight, all the software and tools resulting from this work (like our current software) will be openly available to the research community (see *Resource Sharing Plan*). One important outcome will be open-source software that allows others to carry out simulations of any device of their choice. Our current software that simulates retinal prosthetic device output is available on GitHub, is being used by several research groups, and has guided or is being used to guide development of five devices (2 retinal, 2 cortical, 1 optogenetic).
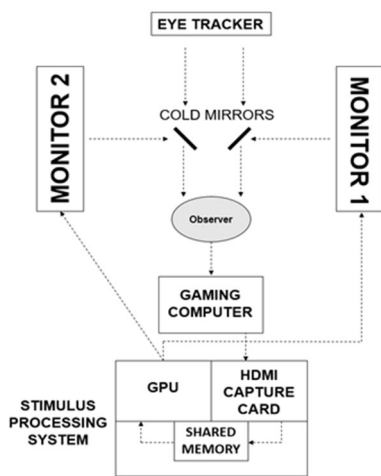
Approach

Stimulus **display**

Figure 2. Schematic of dichoptic stimulus display and real-time filtering.

Video monitors will be the only light source in the room. **Dichoptic.** In Aim 1 we will present dichoptic stimuli (different image to each eye) using a custom-built stereoscope, Figure 2. The device consists of two cold mirrors (used to separate infrared light from non-infrared light, allowing for eye-tracking) mounted by two posts, rotated at a 45 degree angle to capture input from two 32" LED monitors (2560 x 1440 pixel resolution) and reflect the monitor images separately into each eye of the observer. The region of monitor reflected by the cold mirrors is 8.84° high × 8.84° wide, and subtends 768 x 768 pixels, each subtending 0.69 min of visual angle. **Single monitor.** In Aims 2 - 4 participants will directly view a large single monitor (70 x 39 cm) at a viewing distance of 57cm, resulting in a 63 x 37 degree display, with each pixel subtending ~1 min of visual angle.

Real-time gaze-contingent image distortions

The eye-position of the participant will be captured using an EyeLink 1000 Plus at ~2000Hz. Our real-time filtering system, **Figure 2**, was built with the ongoing collaboration of NVIDIA Research (David Luebke). A separate computer runs the video game, controlled via a standard gaming controller. A KONA 4-Channel HDMI Capture Card is used to stream the HDMI output of the gaming computer to a stimulus processing computer at 30 Hz. Onboard the stimulus computer, the Capture Card directly passes each frame to a high-powered NVIDIA Quadro RTX 6000 graphics processing unit (GPU) without the involvement of the central processing unit (CPU). The GPU filters each frame, incorporating eye-position information (Aims 2-4), using the Compute Unified Device Architecture (CUDA), a parallel programming framework developed by NVIDIA, using software developed in-house. This architecture can preserve the 30 Hz stream rate with minimal lag (~1 frame).

Aim 1 distortions are not dependent on eye-position. In *Aims 2* and *3* the distortions due to axonal comets and cortical neuronal architecture depend upon the retinotopic location of the stimulus. In *Aim 4* eye-position information is necessary because the magnitude of the temporal distortions depend on how rapidly the stimulus changes on the retina.

Video Game Perceptual Training

The protocol consists of two main components. Video game training, which is used to try to induce plasticity (which may be preceded or followed by brief visual deprivation), and tests of object recognition which will be carried out by asking participants to detect whether or not an object is present in a scene, see **Table 1.**

1. *Familiarization:* Participants will train on the game Fruit Ninja with undistorted input until they reach proficiency, quantified as being able to slash 60 fruit before the game ends, for three games in a row (game over is triggered by accidentally slashing a bomb or failing to slash three fruit).

2, 3b, 4. Tests of object recognition: We will ask participants to identify whether or not a cued object is present in a dynamic scene. The object dataset currently consists of 17 scenes, and 45 objects. As shown in **Figure 3**, Participants will be cued to report whether or not a specific object is present in the scene. Target object absent trials contain a distractor object. The scene will appear for 2s and is 'panned' (including both an expansion/contraction zoom and a random direction of translational motion) over time with variable speeds to create a dynamic image. Unlike Fruit Ninja, the object will not be moving in relation to the background, forcing the participant to rely on static object recognition cues. Each test trial consists of a
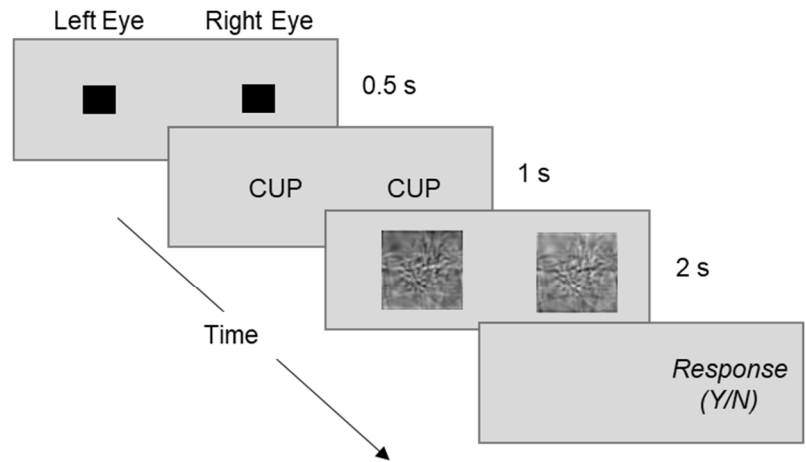

Figure 3. The object recognition task. On each trial the subject reports whether or not the cued object is present.

random object-scene pair (17 scenes, 45 objects, 5 possible object sizes, 5 different zoom speeds, resulting in a total of 19125 possible combinations) Moreover, on each trial both the location of the object in the scene and its angle (+/- 30 degrees) is also varied randomly, making it impossible to learn the task by memorizing exemplars. Hits, correct rejections, false alarms and misses will be converted into measures of d-prime.

Trained distortions: In a pre- and post-training test, and in interleaved test intervals (after every 10 1-hr blocks of video game training) we will measure performance for object recognition with trained distortions. This, along with fruit slashing scores, will track the time-course of learning.

Transfer of learning to novel distortions: One possibility is that plasticity will ameliorate the specific spatiotemporal distortions on which participants have been trained. An alternative (non-exclusive) possibility is that participants will simply get better at extracting information from distorted input. To examine this, at the beginning and end of training we will measure test performance using untrained distortions, described more specifically for each specific Aim.

Table 1. Training and testing protocol

protocol		input	task	duration
1. Familiarization		Undistorted	Fruit Ninja	2 x 1 hr. sessions
2. Pre-test		Undistorted, trained & transfer conditions	Object recognition	3 x 1 hr. sessions
Interleaved	3a. Fruit Ninja distortion training	Trained distortions	Fruit Ninja	3 sets. 10 x 1 hr. sessions
	3b. Intermediate object recognition	Trained distortions	Object recognition	2 sets, 1 x 1 hr. sessions
4. Post-test		Undistorted, trained & transfer conditions	Object recognition	3 x 1 hr. sessions

3a. Video game training. Learning, especially for low level visual features that are normally resistant to generalizable learning, can be enhanced through gamification[79]. Video games enhance visual performance in normally sighted adults across a wide range of tasks[80-82] with effects that seem to be long-lasting; remaining even two years after the end of intervention[83,84]. More recently, video games have begun to be used to treat both adults and children with amblyopia[85-90].

Although traditionally first-person shooter video games are used for perceptual training[91,92] we will use an in-house customized version of Fruit Ninja. Fruit Ninja has all of the qualities thought to elicit generalizable

perceptual learning: *"constantly monitoring the periphery for frequent, widely distributed, unpredictable events that require quick and accurate aiming responses"*[80], it is easy to learn and play (even in grayscale) and requires rapid discrimination between similar objects (something most first person action-based games lack). The game has a simple interface, uses a first-person point of view, requires effective monitoring of the entire field of view, has a unidimensional measure of performance (number of sliced fruit), allows for more fine-grained measurements (e.g. performance as a function of fruit size/speed), and modifiable open-source versions in Unity are available (since participants are playing grayscale with distorted input we will create a version in which the various fruit and the bomb are more distinct in shape).

Once participants are familiar with Fruit Ninja using undistorted input, we will train participants using distorted visual input. Our goal is to carry out 30 hours of distorted Fruit Ninja training per participant. This training will be carried out in 3 sets of 10 one-hour blocks. Performance (the number of fruit slashed in each game) will be recorded.

Across all Aims, pilot simulations confirmed that, for the parameters specified in this proposal, the ability to play Fruit Ninja and perform the object recognition task is limited by the 'neurophysiological distortions" of interest in this proposal, rather than engineering device limits such as electrode size, resolution or field of view – i.e. increasing the number of electrodes did not result in appreciable improvement in percept quality.

Visual Deprivation

An initial series of experiments will be used to finalize whether or not we use deprivation during training in *Specific Aims 1-4*. We will examine whether short-term deprivation, via blindfolding, enhances the effects of video game training using distorted input. We will assign participants to two groups (N=20 for each group). Both groups will undergo 10 hours of video game training using the spatially distorted input of *Specific Aim 2* (axonal distortions). One of the two participant groups will undergo visual deprivation before (120 minutes) and after (120 minutes) each video-gaming training and test session.

Previous studies have suggested that both long term[68-73] and short term (days or hours[55,59,74-76]) deprivation can induce plasticity. However all of these studies focused on adaptations to sensory loss, such as loss of input to one eye. Consequently, we do not yet know whether the effects of deprivation-induced plasticity are limited to homeostatic alterations of neural responsiveness and/or scaling of the population of receptive field sizes influencing perception[93]. Importantly, the distortions of *Specific Aim 2* do not include loss of input and cannot be compensated for via suppression. Thus, perceptual plasticity cannot be mediated by homeostatic regulation of sensitivity and/or a simple increase in the sensitivity or size of receptive fields[93]. *If we see enhanced plasticity with deprivation then this would imply that short term deprivation can facilitate fundamental shifts in how early visual features are decoded.*

If we see stronger learning when participants are deprived before and after training then we will examine whether it is the pre-deprivation, the post-deprivation or both that are critical, and we will utilize the optimal deprivation regime in the experiments of *Specific Aims 1-4*.

Participants

For each experiment we plan to collect data on N = 20 participants (age 20-77). Piloting will be done in undergraduates, but the main experiments will be carried out in individuals aged 50-77.

We do not anticipate difficulties recruiting a participant pool that roughly matches Seattle demographics in regard to race and ethnicity.

Our hope is that Fruit Ninja, despite being a simple game that requires minimal pre-training, will produce similar amounts of perceptual learning to an action RPG. If not, then we will use an action RPG, and modify recruitment accordingly, as described in ***Inclusion of Women and Minorities***, and ***Inclusion over the Lifespan***.

[handwritten: not bold.]

Control group

As is demonstrated by pilot data of **Figure 4**, participants improve with training when repeatedly carrying out the object recognition task (testing was carried out with novel objects). For each Specific Aim we will therefore compare performance to a control group (N = 20) who will carry out with full protocol except 3a (video game training with distorted input).

Statistical analyses

See *Statistical Design and Power*

Aim 1. Abnormal Neuronal Population Responses: can virtual patients learn to interpret signals generated by simultaneous stimulation of on and off cells?

In natural vision, the on and off-cells of the early visual system carry complementary information – regions of the image that produce strong on-center cell activation do not excite off-center cells, and vice versa. However, electronic implants (both retinal and cortical) simultaneously stimulate on- and off-cells indiscriminately (unless the resolution permits for single cell stimulation). In the case of optogenetic proteins, it may be possible to selectively target either on or off-cells, but it would be extremely difficult to target them both in the complementary manner that would mimic natural coding.

It is possible to successfully adapt to _congenitally_ abnormal on and off-cell population responses. Individuals with complete Schubert–Bornschein CSNB1 genetic deficits have severely compromised on-bipolar pathways[94,95]. Yet these patients show surprisingly good visual performance under photopic conditions, with an average visual acuity of 0.3 logMAR (Snellen 20/40)[96] and report no perceptual difficulties beyond their acuity loss (M. Neitz 2015, personal communication).

However, little is known about whether it is possible to learn a novel decoding of on- and off-cells in adulthood. Only one experimental study has specifically examined plasticity in accessing electrically stimulated patterns of activity within V1 in
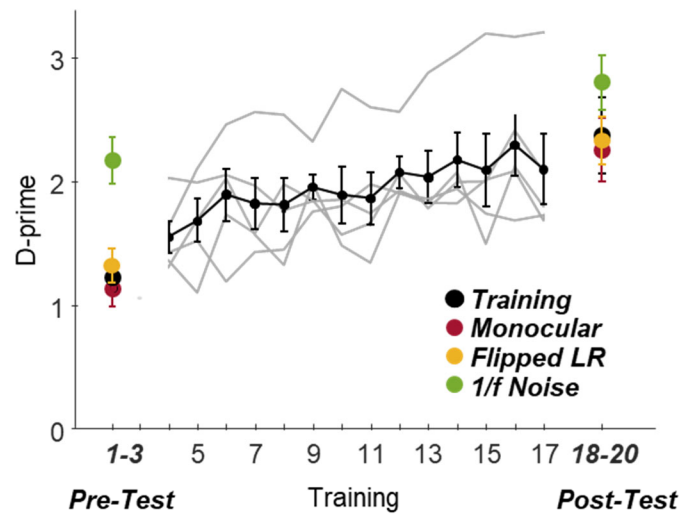

Figure 4. Pilot data. Subjects were pre-tested on the four conditions described in Specific Aim 1 (large solid symbols represent group average data, averaged across three sessions). Subjects then repeatedly carried out the object recognition task, as described in General Methods, with dichoptic Fourier filtered input, as described in Specific Aim 1. Each session lasted ~1 hour/~400 trials. Significant improvements in performance were observed; differences in d' between the first and last session were significant at p<0.0001 for all 5 subjects. Individual subjects are shown with gray lines, black lines and solid symbols represent group averages. Transfer of learning was then measured across all four conditions.

adult animals. Ni and Maunsell[56] examined the effect of prolonged training on detection thresholds for microstimulation within V1 of the macaque. It was possible to train the animals to become experts at detecting V1 microstimulation, but this came at the cost of impaired detection of real visual stimuli at the same retinotopic location. Interestingly, this effect was reversible after retraining on detecting the visual stimulus – suggesting that the local circuitry (whether within V1 itself, or within a higher cortical area decoding V1 activity) was capable of dynamically reconfiguring to better detect task-relevant patterns of neuronal activity.

Experimental design. It is not possible to use normal visual stimulation to simultaneously excite on and off-centered ganglion cells in the retina. Instead, we will use dichoptic stimulation to produce conflicting on and off-cell population responses within binocular cells in the primary visual cortex. We assume that conflicting contrast-reversed images can, within V1, act as a proxy for inappropriate on and off-cell population responses, because regions that would, in the original image, produce strong on-responses will produce strong off-responses in the contrast reversed image, and vice versa. As shown in **Figure 5**, conflicting dichoptic on-off input will be induced using each greyscale image (**I**) and its inverse (**I′**), filtered using a radial checkerboard (**F**) and its inverse (**F′**) in Fourier space. The original (**I**) and its contrast-reversed equivalent (**I′**) will be converted in real-time into the Fourier domain, multiplied by one of the two Fourier filters, and then converted back to image space using the inverse Fourier transform. The middle panels shows the 4 examples of possible filtering: **I * F, I * F′, I$_{cr}$*F, and I$_{cr}$* F′** (where * denotes convolution). Note that the sum, **[I * F] + [I * F′],** equals the original image **I**. The sum, **[I$_{cr}$ * F] + [I$_{cr}$ * F′]**, equals the original contrast reversed image **I$_{cr}$**. In one eye, we will present the sum of two filtered images, **[I * F′] + [I$_{cr}$ * F]**, such that half the spatial frequency and orientation content is based on the original image and the other half is based on the contrast reversed image. In the other eye we will present
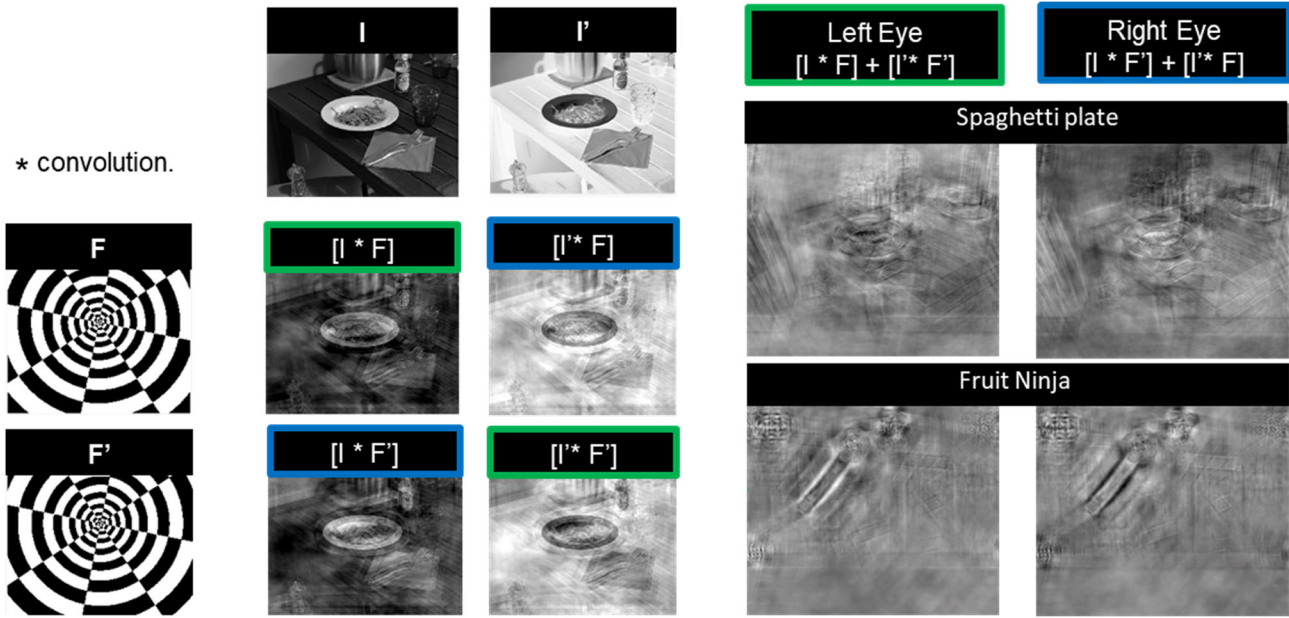
*Figure 5. The upper two panels show an example scene (**I**), and the contrast-reversed version of that scene (**I'**). The leftmost panels show two Fourier filters: **F** and **F'**. Filters are shown in the Fourier domain, with spatial frequency increasing with distance from the center of the image and orientation changing along the polar angle dimension. Each filter is a complement of the other, so the full spatial frequency and orientation content of the scenes is divided equally across the two filters. The lower middle panels show the convolution of the original and contrast reversed images with Fourier filters **F** and **F'**. Rightmost panels show both the plate of spaghetti and the still frame of the Fruit Ninja video of Figure 1, dichoptically filtered, as presented to left and right eyes respectively. Although these images do not resemble the perceptual experience of simultaneous on-off stimulation, interpretation of these images requires an analogous process of suppressing inappropriate contrast-reversed visual input.*

the sum of the complementary filtered images, **[I * F] + [I$_{cr}$ * F']**. *Thus, across both eyes, all the spatial frequency and orientation information of both the original and contrast reversed image is preserved.* However, the sum of the two eyes is a blank image, and the image presented to each eye is heavily degraded (see rightmost panels).

There are of course major differences between our dichoptic paradigm and the effects of electrical stimulation. Critically, in electrical stimulation the unnatural stimulation pattern is elicited monocularly, whereas in our paradigm it is elicited dichoptically. Differences in input between the two eyes are normally interpreted as a depth cue or one eye is suppressed. However, it is believed that depth signals emerge exclusively from same-sign binocular cortical filters[99]. Given that our dichoptic input images are contrast anti-correlated, it is not surprising that our pilot participants never reported perceiving apparent depth.

Using these filters, we will distort the video stream in real time during video game training. Resolution and field of view will be limited by the monitor display [768x768 pixels, each pixel subtending 0.69 min of visual angle, corresponding to a device subtending ~20/14 Snellen acuity; Field of view = 8.84° high × 8.84° wide].

In a preliminary study, we measured improvements in performance with training on an object recognition task: participants were presented with on-off scrambled images containing a scene that had a 50% chance of containing an embedded object or a distractor object, as described above, and were instructed to indicate if they saw the embedded object. Preliminary results show that participants show rapid improvement in identifying objects across sessions, **Figure 4.**

Experiment 1: We will compare performance across 3 conditions:

1. Trained conflicting on-off input. We will measure improvements in d-prime for object recognition with training, using the filters described in **Figure 5**.

*2. Transfer of learning to monocular presentation (red symbols, **Figure 4**).* In our pilot data, performance was very similar for monocular (regardless of eye/filter) as for dichoptic presentation, both before and after training. This suggests that participants are learning to extract on-off scrambled information independently from each eye – rather than binocularly combining information.

*3. Transfer of learning after switching the filters across eyes (yellow symbols, **Figure 4**).* In pilot data, learning transferred fully when filters were switched across the two eyes, suggesting that perceptual learning is not specific to each eye's filter.

We will examine whether participants are in fact suppressing one eye throughout the experiment by setting the images in each monitor to be red or green, and asking participants to report on the perceived color of the scene as well as the identity of the object. If the scene is consistently red or green (rather than yellow) within each trial then this would suggest that participants are performing the task monocularly, and that learning involves suppressing monocular contrast-reversed information. Although the monocular version of this task does not involve directly conflicting on- off stimulation (in which on- and off cells at the same location, with the same spatial frequency and orientation tuning are simultaneously firing), adaptation does require a similar process of learning to suppress a subpopulation of V1 cells, based on their spatial frequency and orientation tuning.

4. Transfer of learning to dichoptic stimulus with 1/f noise (green symbols, **Figure 4***).* In our pilot data, both pre- and post-training performance was better when the contrast reversed information *(I')* was replaced with 1/f noise. With training, the difference between using contrast-reversed and 1/f noise was reduced, showing that training allows participants to suppress monocular contrast-reversed information more effectively.

These preliminary results represent a fundamentally new type of perceptual learning. It has previously been shown that orientation-specific deprivation results in an increase in responsiveness of neurons tuned to that orientation[59,76], suggesting that a reweighting of neuronal populations on the basis of visual experience is possible. One critical difference between the proposed study and these previous studies is that our distortions do not selectively deprive cells of input at a certain spatial frequency or orientation. Consequently, perceptual compensation cannot be mediated by a simple homeostatic increase in responsiveness due to loss of input[93], but rather requires dynamic experiential reweighting on the basis of which cells contain a useful visual signal. It has also been previously shown that perceptual learning can refine participant templates so as to better match a particular visual target[49,98]. However, this learning is thought to involve "reweighting of connections between basic visual channel outputs and a learned categorization structure" (i.e. learning a novel template)[49], rather than being mediated by a reconfiguration of the fundamental building blocks of visual perception.

Aim 2. Systematic Spatial Distortions: can virtual patients learn to compensate for the spatial blur induced by axonal stimulation?

Every retinal ganglion cell has an axon that traverses the retinal surface en route to the optic nerve. Animal models[29,31-33,100], modelling[34] and human data[35] suggest that retinal implants produce significant axonal stimulation, which results in elongated percepts that resemble 'comets'. Comet elongation depends on the relative ratio of ganglion axonal to ganglion/bipolar soma stimulation which is primarily determined by (1) retinal degeneration, (2) distance of the electrode from these different cell types, and (3) pulse duration (longer pulse durations advantage bipolar cells). Distortions are most severe in epiretinal implants (that lie between the ganglion axons and the vitreous)[14], but are also observed in patients with subretinal implants (that lie between bipolar cells and the choroid)[101].

Because these distortions are mediated by axonal tract pathways, they do not represent a simple convolution of the image with a stationary
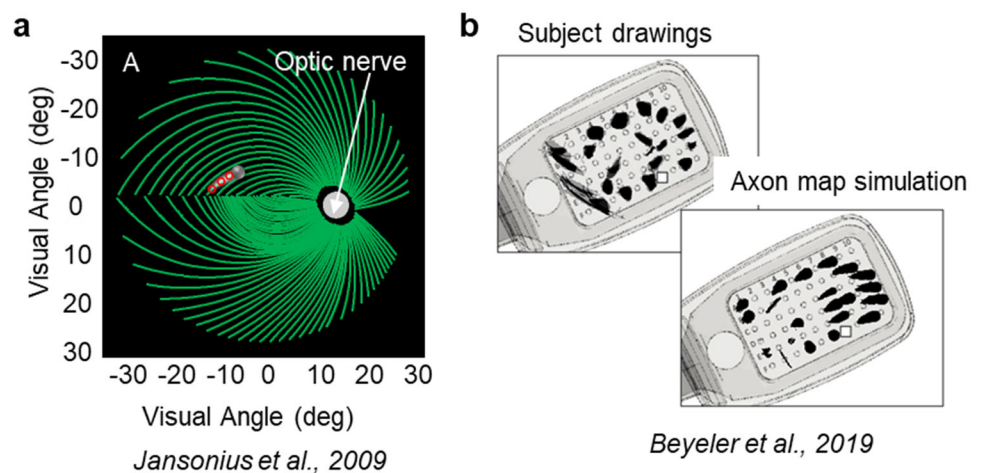

Figure 6. Spatial distortion due to stimulation of retinal axonal pathways (a) The retinal image is flipped so that the upper region of the retina represents the upper visual field. Green lines represent modelled axon trajectories (based on a computational model of axon fiber trajectories developed using traced nerve fiber bundle trajectories extracted from fundus photographs of 55 human eyes[9]), red circles represent ganglion cell bodies whose axon fibers pass underneath the electrode on their way to the optic nerve, yellow shading represents a perceptual 'axon-comets'. (b) Patient phosphene drawings and cross-validated phosphene predictions of the axon map model, overlaid over a schematic of each subject's implant. A comparison of log mean prediction error shows that the axonal model predicts subject phosphene drawings better than the scoreboard model for each of the four patients[14]. As shown in **Video 1**, this results in streaking that varies across the visual field – as a result the shape of objects change as they move across the retina.

filter (e.g. warping, blurring or translation of retinotopic space), but rather the distortion varies smoothly across the retina in a complex way (though with a discontinuity near the horizontal Raphe), see **Figure 6, Video 1**.

It is unlikely that these distortions can be overcome by reorganization of the retinotopic map in V1. Overall, the literature suggests that the retinotopic organization of early visual areas (retina-V2) is primarily determined by molecular signaling, with visual experience playing the role of refining the map[102-112]. There is limited plasticity within retinotopic organization, even when the disrupted input is congenital. For example, rod monochromats lack cone photoreceptor function and therefore have a congenital retinal scotoma within the all-cone foveola. Although some limited reorganization can be observed[71], it is likely to be mediated by an expansion of receptive fields rather than by a fundamental reorganization of the retinotopic map. Individuals with miswiring of the retinal-fugal projection (achiasmatia[113], albinism[114,115], FHONDA syndrome[116], or individuals born with one cortical hemisphere[117]) do not show compensatory plasticity: instead these individuals have overlapped cortical representations of left and right visual hemifields, wherein each region of cortex represents two distant (mirror symmetric) locations in visual space[114,115].

As far as adult plasticity is concerned, the overall literature suggests V1 retinotopic organization is relatively immutable. Across several electrophysiology studies Gilbert et al. found that neurons in the lesion projection zone of an adult-onset scotoma rapidly became responsive to visual stimulation of the retina surrounding the damaged area[69,118-120]. However, studies across multiple other laboratories have failed to find evidence of cortical reorganization after retinal lesions[68,121,122]. Similarly, while some human fMRI studies have shown responses in the LPZ[123-125] or shifts in measured cortical retinotopic organization[126] in late blind individuals suffering from visual field loss due to retinal dystrophies, other studies have failed to find evidence for V1 reorganization[72,73,127]. All of the discrepancies between reported outcomes in both the neurophysiology and the fMRI literature can be explained by the presence of one of two possible confounds in those studies that *did* find plasticity: (1) using a model to estimate cortical retinotopic organization that does not consider the absence of input when the stimulus was in the scotoma[128], and/or (2) failing to account for top-down attentional effects[73,127]. (Also see[129,68], for additional potential confounds).

Although the balance of evidence suggests that the neural connection pattern underlying retinotopic maps in V1 seems to be relatively immutable, there does seem to be considerable plasticity in the ability to decode congenitally distorted V1 signals. Despite their dramatically abnormal retinotopic representation, acuity losses in achiasmatic and individuals with albinism seems to be primarily due to their foveal hypoplasia and there is no perceptual confusion across mirror symmetric locations in the two visual hemifields[113]; suggesting that the abnormal retinotopic organization within V1 is successfully perceptually decoded by later stages of visual processing.

Finally, the most relevant example of successful decoding of congenital retinotopic distortions comes from PD, an individual who had congenital central cataracts resulting in annular pupils until the age of 43. PD showed a host of perceptual adaptations to his distorted visual input, including suppression of diplopic images[130] – an example of spatially specific context-dependent retinotopic suppression very similar to what would be necessary to compensate for distortions associated with axonal stimulation.

The capacity of the adult visual system to successfully decode novel distortions in the retinal input is still relatively understudied. Some evidence for perceptual compensation comes from individuals with macular degeneration: shapes presented adjacent to the scotoma are distorted – elongated in the direction of the scotoma[131]. Curiously, analogous distortions can be generated in visually typical individuals in the blind spot – and they appear surprisingly quickly – after mere seconds[132], suggesting these distortions might be mediated by short term homeostatic regulation[93]. Another classic example is of course prism adaptation, however this is generally considered to be primarily mediated by short-term visuo-motor adaptation[133]. Finally, retinally specific decoding of retinotopic distortions in adulthood over a time course of a few weeks can be observed in almost every individual over the age of 50 in high-income countries: successful adaptation to ever more complex and expensive spectacle or contact lens prescriptions.

Experimental design. We will distort the video stream in real time using gaze-contingent updating of distortions. We will simulate the relative proportion of axonal and somal stimulation expected from a sub-retinal array in non-degenerate retina[97] using a model previously validated in Argus II patients[14]. We selected parameter values describing electrical and neural spread of activation parallel to (λ = 250 microns) and orthogonal to (ρ = 100 microns) axon pathways. For comparison the size of the average ganglion soma is ~8 microns[134]. These parameters are smaller than those observed in Argus II patients, but are plausible for an 'idealized' subretinal

array with small electrodes. [50x89 electrodes, radius = 42 microns/0.15 deg of visual angle. Center-to-center electrode spacing = 169 microns/0.59 deg of visual angle, corresponding to a device subtending Snellen acuity of 20/700; Field of view = 30° high × 52° wide; biphasic pulse trains; pulse-width = 1ms]. Although this resolution is low, axonal streaks are still the limiting factor for image quality. Participants will be asked to do the object recognition task, using distorted input, as described in *General Methods*.

Experiment 2-1: We will compare performance across 3 randomly interleaved conditions, in both pre- and post-test.

1. Trained distortions. We will see to what extent participants improve in their ability to compensate for simulated axonal distortions with video game training.

2. Rotated distortions. One possibility is that plasticity will ameliorate distortions along the specific axonal pathways for which participants have been trained. An alternative (non-exclusive) possibility is that participants will simply get better at extracting information from distorted input. To examine this, in pre- and post-training tests we will measure performance using distortions based on axon map pathways rotated by 90, 180 and 270 degrees. Significant improvements in performance that do not transfer to rotated axon maps would imply that participants have learned to compensate for specific spatial distortions that do not represent a simple warping or translation of retinotopic space, but rather vary across the retina in a highly complex way.

Experiment 2-2: As described above, Dilks and others have shown perceptual compensation for distortions in the retinotopic map within individuals with macular degeneration and other visual lesions (including the blind spot[132]): shapes presented adjacent to the scotoma are distorted – elongated in the direction of the scotoma[131]. One critical difference between the proposed study and previous studies by Dilks is that axonal comets do not result in a scotoma. Therefore, perceptual compensation cannot be mediated by a simple expansion of receptive fields due to loss of input[93], but rather requires a genuine recalibration of the spatial decoding of V1 input to compensate for induced distortions. We will use the axonal stimulation model to simulate the distortions of single circular or elongated (oval) spots of light. Before and after training, on each trial participants will be presented with a shape (oval or circle), blurred based on our axonal model, in two different locations of the visual field. They will be asked to report whether the stimulus is a circle or an oval. If participants are simply learning to recognize objects within a distorted world, then performance on the task should not change with training. In contrast, if individuals are perceptually adapting to the specific axonal distortions to which they have been trained, in a retinotopically specific manner, then participants will become less sensitive to elongations along axonal pathways. This work will be synergistic with the R00 phase (at UCSB) of a recent K99 awarded to Michael Beyeler in Dr. Fine's laboratory, which includes the Aim of further refining a computational model that predicts the perceptual distortions of retinal prosthesis patients.

Aim 3. Abnormal Cortical Neuronal Population Responses: can virtual patients learn to interpret distorted signals induced by stimulating the V1 neural architecture?

Stimulating the cortex to produce interpretable percepts provides its own set of unique benefits and costs. Certainly, the massive expansion of the foveal representation in V1 could allow for relatively high sampling of spatial position, as compared to retinal stimulation. However, V1 receptive field sizes are relatively large, even in the fovea - around 1 degree of visual angle. The large size of V1 receptive field sizes in the fovea is somewhat counterintuitive, given the extremely high resolution of Vernier (0.3-1 min arc[135] and grating acuity ~60 cycles/degree). Our ability to perceive a single point of light, a fine grating, or the offset in a thin line relies on a complex pattern of responses across a population of neurons with center/surround receptive fields. (An insight from Fourier analysis, that may or may not help, is that a discrete point of light contains an infinitely broad range of spatial frequencies; thus, if early visual areas are carrying out a wavelet analysis one would expect the resolution of small spots of light to be mediated by population responses across nearly all V1 neurons with receptive fields covering that spot).

As shown in **Figure 7**, simulations of the effects of electrically stimulating the cortical architecture suggest that, for a fixed electrode size, phosphene size increases (roughly linearly) as a function of eccentricity, due to receptive field sizes increasing as a function of eccentricity[16,17]. At the fovea this linear slope flattens to a lower limit where phosphene sizes are ~1.4 degree diameter, regardless of electrode size. As a consequence, our simulations predict no appreciable difference in phosphene sizes for electrodes ranging between 0.01-1 degree micron radii, regardless of eccentricity, suggesting that neurophysiological rather than engineering constraints are likely to limit the spatial resolution of cortical prostheses.

Orientation and ocular dominance columns are relatively large: >2mm for a full ocular dominance/pinwheel map[16,28]. This means that stimulation with extremely small electrodes is likely to have no benefit in terms of reducing phosphene size, but could potentially stimulate specific ocular dominance and orientation columns, with the potential of creating percepts that are elongated and potentially binocularly rivalrous.

In normal vision, we are exquisitely adapted to eccentricity-dependent levels of blur – our perceptual world appears universally sharp despite a steep decline in high frequency information (due to optical, retinal and cortical factors) as a function of eccentricity. Individual differences are large, and there is evidence that individuals are calibrated to their own eccentricity-blur function[44-46,136]. Our cortical prosthetic essentially provides a new (and worse) function describing blur as a function of eccentricity. We know that individuals can adapt to blur[137], but it is not yet clear whether individuals can learn to adapt to blur in an eccentricity-dependent manner, as occurs in normal vision.

Experimental design. We will distort the video stream in real time, using gaze-contingent updating of distortions. We will simulate a depth electrode array that equally stimulates cell bodies (but not axons) across all cortical layers. Our electrode sizes are chosen to minimize percept elongation/binocular rivalry while maximizing resolution, and our electrode spacing across cortex was chosen to result in even coverage across the visual field [98x124 electrodes, electrode radius = 10 micron, Center-to-center electrode spacing = 0.5 degrees of visual angle, projected onto the cortical surface, corresponding to a device subtending ~20/600 Snellen acuity; Field of view = 48° high × 60° wide; biphasic pulse trains; pulse-width = 1ms].

Experiment 3: *bold, underline.*

We will begin by determining the degree to which our virtual patients learn to compensate for the eccentricity dependent distortions associated with cortical stimulation.

We will then examine whether individuals have learned to adapt to blur in an eccentricity-dependent manner, as occurs in normal vision. Before and after training, on each trial participants will be presented with two images of the same object, blurred based on our cortical implant model, in two different locations of the visual field, and will be asked to report which image seems more blurred[137]. Object images will be blurred using the function $\sigma = s * e + i$, where σ is the standard deviation of the phosphene Gaussian along the minor axis, e is eccentricity, $s = 0.05$, $i = 0.69$, parameters based on[17], as shown in **Figure 7.** During training we will always use $s = 0.05$[17], but during testing we will vary s to find the value that best predicts subjective blur equality across the entire visual field. Before training stimuli should appear equally blurred when $s = 0$ (phosphene sizes of ~0.77 degree radius across the entire visual field), regardless of eccentricity, since this value represents equal blurring across the visual field, and people have already adapted for their own underlying visual systems. After training, if participants have fully adapted to their cortical implant, then the slope that best predicts subjective equality should match the cortical magnification function for the trained cortical implant.

move

Aim 4: Abnormal Temporal Kinetics: can virtual patients learn to adapt to the slow offset kinetics of optogenetic proteins?
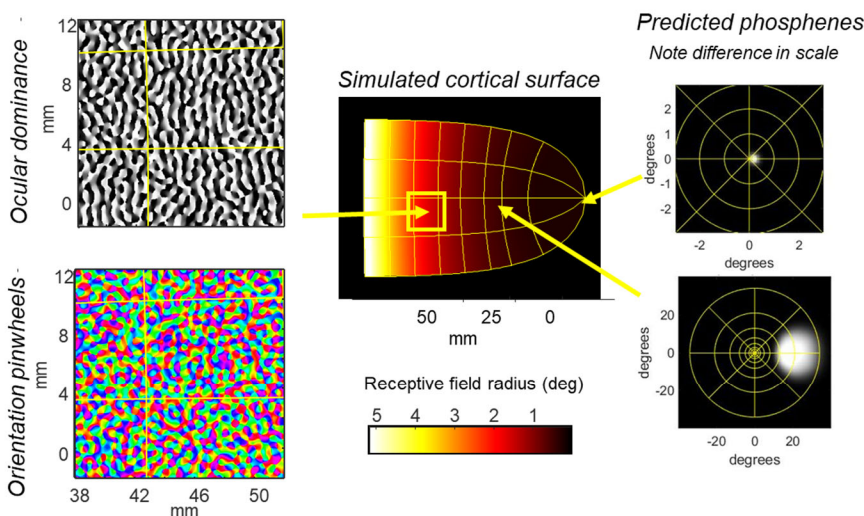

*Figure 7. Simulation of the perceptual distortions induced by cortical stimulation. Center panel shows simulated V1 cortical surface [15] showing receptive field sizes[16,17]. Leftmost inserts show simulated ocular dominance [28] and orientation pinwheel maps. We simulate cells as orientated Gaussians, ignoring the complications resulting from simultaneously stimulating on and off-cells since this is dealt with in Aim 1. The neural signal resulting from stimulation of this cortical surface can be represented as a predicted percept. Rightmost panels show phosphene size for a very small electrode (0.10μm radius) for two electrode locations, note change of scale. The same electrode size produces radically different phosphenes, depending on eccentricity. Two example phosphenes are shown. As shown in **Video 1**, this results in eccentricity dependent blurring across the visual field.*

Very little is known about the ability of the human visual system to adapt to abnormal temporal dynamics, though adaptation to temporal blur has been observed[138]. There is evidence for compensation for the different temporal dynamics of rod and cone vision, but it is not known whether this compensation is 'hard-wired' or driven by experience[139]. There is also evidence for changes in the temporal dynamics of visual processing with aging[140] [140,141], but it is not yet clear whether or not compensatory plasticity occurs as a result.

Currently, both optogenetic proteins and small molecule photoswitches tend to have relatively slow kinetics, and it is not clear whether these kinetics can be improved without either loss of sensitivity or using higher wavelength regions of the spectrum which can cause retinal damage at high intensities. Although more sensitive optogenetic proteins with faster dynamics are being developed[142,143], they still have relatively slow dynamics and/or low sensitivity compared with normal photoreceptor responses[144,145]. Similarly, while small molecule photoswitches have a response to the onset of light that is faster than that of normal photoreceptors (likely due to the lack of phototransduction), the return to ground state for current remains relatively slow even in the most recent more rapid photoswitches[146].

Slow dynamics do not simply reduce the ability to process rapidly moving objects. Rather, they lead to motion streaks (similar to the blur seen in a moving cursor using a sluggish monitor) and thereby reduce the contrast of moving objects, especially larger objects. Critically, this streaking is induced by all forms of retinal motion, including those induced by eye (or camera) movements. **Figure 8** shows a simulation of the expected perceptual effects of sluggish photokinetics, using the dynamics of MW-Opsin used optogenetically[147]. **Figure 8b, c** show the central frames of a movie showing the perceptual effect of these sluggish temporal dynamics, also see **Video 1**. Stationary objects in the scene, such as the fence, are almost unaffected. However, there are significant motion streaks in the moving ball, causing it to, rather dramatically, almost completely disappear. It can easily be appreciated that the disappearance of rapidly moving objects (such as cars!) raises significant safety concerns.

Experiment 4: Using the video game protocol outlined above, we will distort the video stream in real time, simulating the time course of MW-Opsin used optogenetically.

Spatial resolution and field of view will be limited by the monitor display [63 x 37 degree display, each pixel subtending ~1 min of visual angle, corresponding to a device subtending Snellen acuity of 20/20; rising time-constant t = 100 msec, decaying time constant t=200 msec, delay of 50 msec].

After video game training with temporally distorted stimuli we will once again test visual performance on our object recognition task. We will measure performance as a function of the retinal spatiotemporal motion energy produced by the scene (dependent on eye-movements, the spatial frequency content of the scene and the pan speed/zoom) and the spatiotemporal motion energy of the object (which also differ in their spatial frequency content, since they are presented at 5 different sizes). Performance should be particularly poor when there are large amounts of motion energy at low spatial frequencies and high temporal frequencies - *counterintuitively this predicts worse performance for larger objects, especially for faster panning speeds*. We can also test this prediction in the Fruit Ninja task by measuring fruit slicing performance as a function of the size and speed of the flying fruit.

We will then examine transfer to untrained temporal distortions by measuring performance using a temporally reversed filter – i.e. a filter with slow onset and fast offset. Improvements in performance that are specific for the trained temporal filter would imply that participants can learn to compensate for specific temporal distortions.

Figure 8. The perceptual effects of sluggish response kinetics. (a) Black: normalized ganglion cell firing rate in an rd1 mouse retina with genetic introduction of MW-Opsin to a short (25ms) light (535 nm, 0.2 mW/cm²) presented at time t=0 (From Berry et al., 2019, figure 2f). Red: our simple model of the neural time-course that uses a rising time-constant t = 100 msec, a decaying time constant t=200 msec and a delay of 50 msec. (b, c) Frames from a movie of a child kicking a football filtered using the simulated temporal dynamics of the MW-opsin (also see **Video 1**). The sluggish temporal dynamics blurs moving objects, affecting object recognition more than static object detection. While the static fence remains visible, the moving ball disappears.